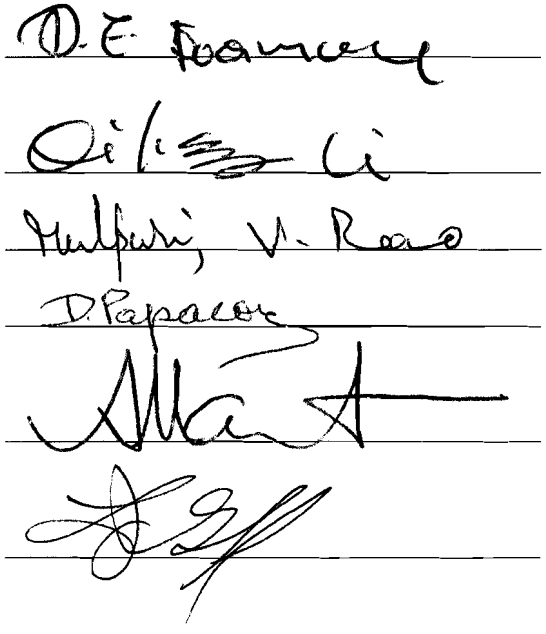
ISSUES OF ESD PROTECTION IN NANO-SCALE CMOS

by

Yang Yang
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Electrical and Computer Engineering

Committee:



Dr. Dimitris E. Ioannou,
Dissertation Director

Dr. Qiliang Li, Committee Member

Dr. Rao V. Mulpuri, Committee Member

Dr. Dimitrios A. Papaconstantopoulos,
Committee Member

Dr. Andre Manitius, Chair, Electrical
and Computer Engineering Department

Dr. Lloyd J. Griffiths,
Dean, The Volgenau School of Information
Technology and Engineering

Date: June 28th, 2010

Summer 2010
George Mason University
Fairfax, VA

Issues of ESD protection in nano-scale CMOS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Yang Yang
Master of Science
Queen's University Belfast, 2005
Bachelor of Science
Beihang University, 2004

Director: Dr. Dimitris E. Ioannou, Professor
Department of Electrical and Computer Engineering

Summer 2010
George Mason University
Fairfax, VA

Dedication

I dedicate this dissertation to my parents and my fiancee, Juan Wang, for their patience, tolerance and constant support during the completion of this dissertation.

Acknowledgments

This work could not have been completed without the support of several people.

First, I would like to express my deepest gratitude to my advisor, Prof. Dimitris E. Ioannou, for his insight, guidance, encouragement and support during my graduate study at George Mason University. His critical thinking skills and clarity of thinking made working with him an enjoyable and rewarding experience. I consider myself extremely fortunate to have Prof. Ioannou as my academic advisor.

I would also like to express my thanks to Akram A. Salman who was an ESD engineer with Advanced Micro Devices, for his guidance and suggestions on the project of field-effect-diode at the initial stage of my graduate study. His continued encouragement and motivation during that project helped me a lot.

Most of the experimental work related to this thesis was done at the IBM Microelectronics Div. Semiconductor Research & Development Center, Essex Junction, VT-05452 while on a student internship. I would like to express my sincere thanks to my manager at IBM, Robert J. Gauthier Jr. and other colleagues, Junjun Li, Kiran Chatty, Souvick Mitra, Rahul Mishra, James Di Sarro, Ralph Halbach and Christopher Seguin for their support and insightful discussions.

My special thanks to Prof. Qiliang Li, Prof. Rao V. Mulpri and Prof. Dimitrios A. Papaconstantopoulos for accepting to serve in my doctoral advisory committee and the valuable knowledge I learned from them during the courses they taught.

The financial support of my research by National Science Foundation is gratefully acknowledged.

Finally, I would like to thank my parents and my fiancee Juan Wang, whose support and love made this work possible.

Table of Contents

List of Tables

List of Figures

Abstract

ISSUES OF ESD PROTECTION IN NANO-SCALE CMOS

Yang Yang, PhD

George Mason University, 2010

Dissertation Director: Dr. Dimitris E. Ioannou

Aggressive downsizing of individual transistors continues to improve the performance of integrated circuits. However, as the transistors get smaller, they become more vulnerable to damage by high current and high voltage Electrostatic Discharges (ESD). As technology scales down, among other things, new materials such as high-k gate dielectrics are incorporated into the modern chip fabrication technologies and Silicon-on-Insulator (SOI) technology is gaining acceptance. These technology advances make ESD protection of silicon chips ever more necessary and challenging. Consequently, the present dissertation focuses on ESD related issues in nano-scale CMOS technologies.

The thesis begins with the investigation of high-k gate dielectric breakdown under ESD-like stress. The stress configuration for transistors in the input receiver will be considered first. It is confirmed that high-k gate oxide breakdown is catastrophic under ESD-like stress. Using the constant voltage stress (CVS) method, the gate oxide breakdown voltages (V_{BD}) of NMSOFETs and PMOSFETs are compared under different stress polarities, in order to identify the worst case scenario. The results are also compared with SiON gate dielectric devices. Next, high-k gate breakdown in the

output driver is explored. The results imply that the input receiver is more susceptible (than output driver) to failure due to ESD induced gate dielectric breakdown. Measurement results also show that V_{BD} obtained by the transmission line pulsing method (TLP) is only slightly smaller than that obtained by the CVS method. Methodologies to improve the breakdown immunity are then proposed with the support of experimental results.

The dissertation then focuses on the degradation of NMOSFETs with high-k gate under non-destructive ESD-like stress. For the stress configuration emulating the output driver, little degradation was observed until the device failed by drain-to-source filamentation. By contrast, for the stress configuration emulating the input receiver, degradations of threshold voltage (V_t), drain saturation current (I_{dsat}) and Si/gate oxide interface were observed. The degradations increase with the effective gate oxide thickness and are more severe under positive stress polarity. Different from Positive Bias Temperature Instability (PBTI) stress, the threshold voltage shift depends on temperature rather weakly, indicating a new dominant charge trapping mechanism active on the time scale of ESD events. These results are then compared with those obtained for transistors with SiON gate dielectric. In addition to V_t, I_{dsat} and interface degradation, the impact of the stress on the gate leakage current and on the subsequent PBTI degradation kinetics is also studied.

Finally, the dissertation presents a thorough investigation of the field effect diode (FED) with the aim to explore its potential for ESD protection applications in SOI technology. It is shown that the doping type and concentration under the two gates has an important impact on the device operation. By careful sizing and doping, FED devices with reasonable breakdown voltage values can be achieved at gate voltage values compatible with the latest technology.

Chapter 1: Introduction

With the concept brought up in 1930 [14] and later put into practice using Si-SiO_2 by Kahng and Atala in 1960 [15], Metal-Oxide-Semiconductor Field Effect Transistor (MOSFET) has been the workhorse of the microelectronic industry for the past three decades. The ever-increasing demand for faster and multi-functional integrated circuits (ICs) has led to significant advances in semiconductor manufacturing techniques. Microprocessor clock frequencies have been increasing from about 2 MHz (Intel 8080), made in 1974, to over 2 GHz today (2010) with the smallest feature size on the order tens of nanometers. The number of transistors in ICs has been increased from a few thousand to more than one billion today as shown in Fig. 1.1.

The advances of the microelectronic industry have been achieved mainly by the persistent downsizing of transistors. Decreasing the feature size of the individual transistor not only improves the overall circuit performance (improvements in density, speed and power consumption) but also reduces the manufacturing cost [1, 16, 17]. However, the reduction in MOSFET channel length leads to the undesirable short channel effects, which cause a reduction in the threshold voltage at which the device turns on, especially at high drain voltages [2]. To keep proper function of the transistors, the scaling of transistors should follow certain rules. The common scaling rules for important transistor parameters and their technological or physical constraints are listed in Table 1.1 [1].

Dennard *et al.* proposed the constant-field scaling rule in 1974 [18] (as shown in the second column of Table 1.1) that one can keep short-channel effects under control by scaling down the vertical dimensions (gate insulator thickness, junction depth,

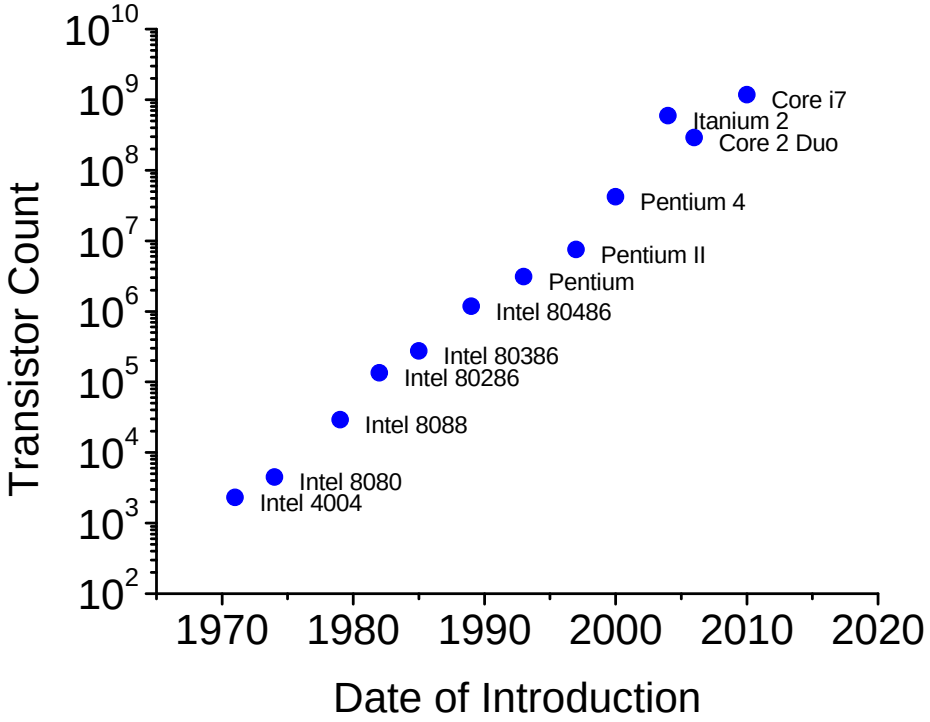


Figure 1.1: Plot of transistor counts against dates of introduction [4].

etc) along with the horizontal dimensions, while also proportionally decreasing the applied voltages and increasing the substrate doping concentration (decreasing the depletion width). As all the voltages and dimensions are reduced by the same factor of k, the electric field in the small-sized devices will remain the same as the large devices. This ensures that the reliability of the small device is not worse than the large one. However, some nonscaling factors prohibit reducing the power supply voltage by the same factor as the device dimension. The primary nonscaling factors include the thermal voltage kT/q and the silicon band gap E_g [2]. The former factor prohibits the scaling of subthreshold current and leads to the non-scalability of the threshold voltage, while the constant silicon bandgap results in the non-scalability of the junction built-in potential, depletion layer width and the short channel effects [2]. Table 1.2 lists the power supply voltage and the device dimensions for several

Table 1.1: A general scaling rule leading to an integrated circuit with a faster speed, and enhanced functionalities have been the major success of the past history of microelectronics [1]

Parameter	Constant field scaling factor	Generalized scaling factor	Limiting factors
Voltage, V_{DD}	$1/k$	ϵ/k	Thermal voltage, quantum confinement
Electric field	1	ϵ	
Channel length L_G	$1/k$	$1/k$	Lithography accuracy
Drain current I_D	$1/k$	ϵ/k	Punchthrough, tunneling between source and drain
Gate capacitance C_{ox}	k	k	Physical thickness limit, gate leakage current
Intrinsic delay CV/I_D	$1/k$	ϵ/k	Non-scalable V_{DD}
System clock frequency f_c	k	k	Parasitic capacitance Interconnect R and C
Chip area	1	1	Yield, delay
# of transistors per chip n	k^2	k^2	Interconnect complexity, yield
Power density $P \propto f_c n C_{ox} V_{DD}^2$	1	ϵ^2	Gate leakage, frequency # of transistors, over heating

technology nodes, which clearly shows the less drastic scaling of supply voltage and the increasing oxide field.

As the voltage cannot scale as fast as the device dimensions, a more generalized scaling rule was proposed by Baccarani *et al.* [19] as shown in the third column of Table 1.1. In this rule, it is proposed that the lateral and the vertical electric fields should be scaled by the same factor so that the shape of the field pattern is preserved. This keeps the two dimensional effects, such as short-channel effects, under control. However, the higher electric field in aggressively scaled MOSFETs does cause serious reliability concerns. For every technology node to be successful, a thorough reliability analysis is required. Within the scope of these requirements, the motivation for this thesis is given in the following section.

Table 1.2: Scaling of power supply voltage, oxide thickness and oxide field in several CMOS technology generations [2]. * this data is taken from the latest 32 nm technology.

Feature size (μm)	Power supply voltage (V)	Gate oxide thickness (Å)	Oxide field (MV/cm)
2	5	350	1.4
1.2	5	250	2.0
0.8	5	180	2.8
0.5	3.3	120	2.8
0.35	3.3	100	3.3
0.25	2.5	70	3.6
0.1	1.5	30	5.0
0.032*	1.0	14	7.1

1.1 Motivation

Gate oxide thickness is reduced for every technology node to maintain the necessary gate capacitance so that the drive current is adequate for improved circuit operation. This gate insulator which is used to block channel current from gate electrodes has already approached the range of 1-2 nm. Besides the huge increase in the gate tunneling current due to gate oxide thinning, the increased gate oxide electric field also poses serious concerns on the reliability of these ultra-thin gate insulators. One aspect of the reliability concern is related to the damage caused by the high-current, high-voltage electrostatic discharge (ESD) events.

Integrated circuits are protected against ESD events through protection networks at each pin of the chip. The protection network should remain in the off-state (with high resistance) during the normal operation conditions and be triggered into the on-state (with low resistance) during ESD events to clamp the pin voltage low enough and avoid damage to the internal circuitry. The normal operation region of the IC and the gate oxide breakdown voltage (and junction breakdown voltage) define the lower and upper bounds of the operation window of the ESD protection network respectively. Fig. 1.2 [5] shows the evolution of this design window as the technology

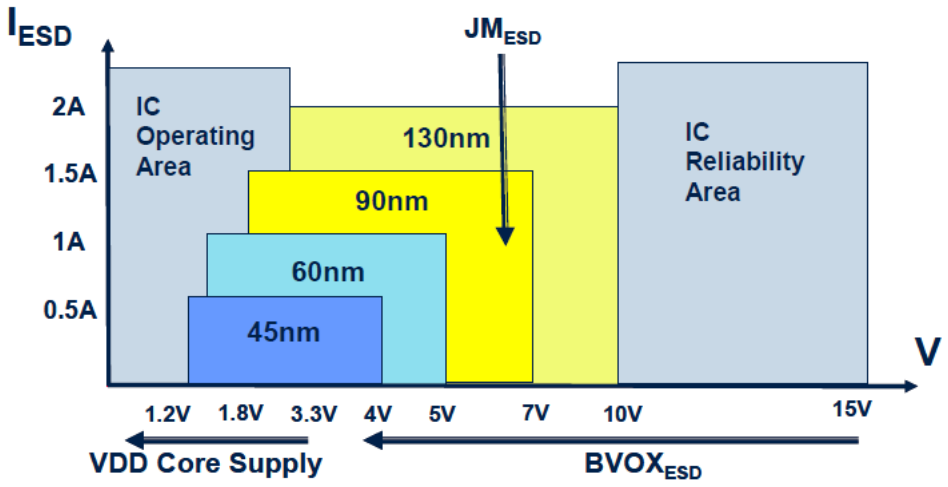


Figure 1.2: Impact of technology scaling on ESD design window [5].

scales down. Although the power supply voltage is decreasing with the technology, the oxide breakdown voltage decreases far more drastically with the scaling of thickness, as shown in Fig. 1.3 which are taken from 45 nm and 65 nm technology. As a result, the lateral dimension of the design window is significantly reduced with scaling.

The shrinking design window challenges the ESD protection robustness and significantly increases the importance of understanding gate dielectric breakdown under ESD-like stress. Especially for charged device model (CDM)-type ESD events, the current can be even higher than 10 A and consequently the IR drop will cause a large voltage drop and overstress the gate oxide of transistors in the input/output (I/O) circuits.

Further complicating matters, new materials with high permittivities (called high-k materials) are introduced to replace the conventional materials such as SiO_2 and SiON to make the gate dielectric. The thickness of SiO_2 or SiON gate dielectric has already approached its physical limit. Further scaling will result in excessively large quantum mechanical tunneling current, which increases the standby power consumption to an intolerable level. High-k materials are helpful to reduce the leakage current

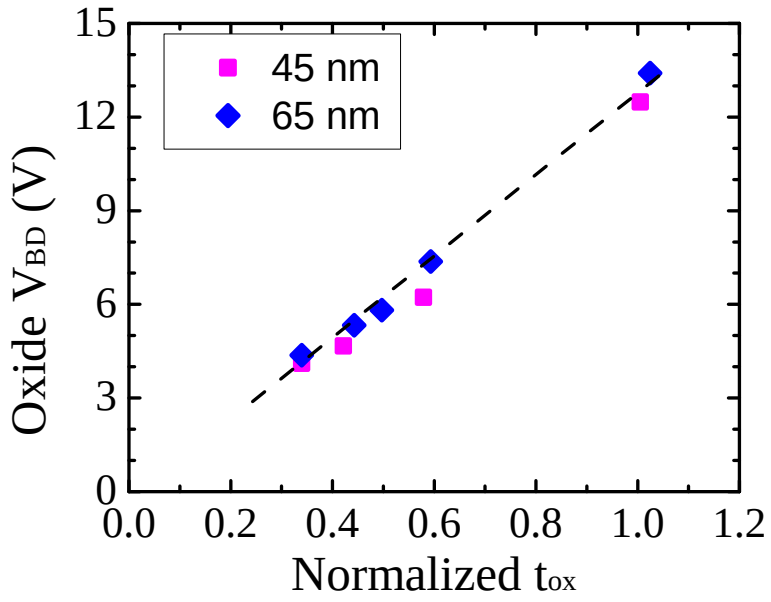


Figure 1.3: Reduction of gate oxide breakdown voltage as the gate dielectric thickness scales down. The data correspond to a breakdown time of 100 ns.

while keeping the benefits brought by scaling [1]. The impact of the new gate dielectric materials on the ESD design robustness is still an open question.

One way to mitigate the above problem is to improve the current protection device or design new device structures which offer more effective ESD protection. This is especially challenging for Silicon-On-Insulator (SOI) technology. The buried oxide (BOX) layer between the active device and the silicon substrate provides the advantages such as low parasitic capacitance, no body effect and high immunity to soft error [2]. However,the BOX layer leads to increased self heating due to its low thermal conductivity and precludes simple integration of common bulk ESD protection devices relying on vertical junctions. For example, silicon-controlled-rectifier (SCR) is not easily feasible in SOI technology. The Lateral diode has been widely used to construct P-N junctions in the silicon layer above the BOX [20]; but it is not capable of providing

[1]The details will be described in Sec. 2.4.1.

local clamping [2]. Although silicide blocked gate-grounded NMOSFET is able to provide local clamping, it has a relatively low failure current. It is therefore desirable to design a new device structure which can provide robust local protection in the SOI technology.

The present thesis investigates high-k gate dielectric breakdown as well as degradation under ESD-like stress. The study is essential for robust ESD design of the latest technology. The results are compared with the conventional SiO_2 and SiON gate dielectric and the impact of transitioning to high-k gate dielectric on the ESD-design robustness are evaluated. For efficient and accurate characterization of gate oxide failure under ESD-like stress, different test methods are compared as well. Based on the experimental results, methods to increase the immunity to oxide breakdown in the I/O circuits are proposed.

To achieve robust ESD design in SOI technology, this thesis also explores a new ESD protection device called Field Effect Diode (FED) [21]. The original FED proposed in [21] does not have a large enough anode breakdown voltage (V_{FB}) under the bias voltage compatible with power supply voltage in the latest technology. As a result, the protection device may turn on during normal operation conditions. For real application, the anode breakdown voltage has to be increased. Based on Technology Computer Aided Design (TCAD) simulation results, the physics of device operation is investigated. A method to increase the anode breakdown voltage is proposed, which makes its applications for ESD protection feasible in the SOI technology.

[2]The concept of local clamping and non-local clamping will be introduced in Sec. 2.3.1

1.2 Dissertation Organization

The dissertation is organized as follows:

Chapter 2 presents the preliminary information, including ESD and gate dielectric degradation and breakdown, for a better understanding of the dissertation. The chapter presents a brief introduction on the ESD phenomena, followed by the description of various ESD models, characterization methods and different ESD protection strategies. This chapter also points out the necessity of introducing high-k gate dielectric, covers the basic physics of gate dielectric degradation and breakdown and reviews the literature results about gate dielectric breakdown and degradation under ESD-like stress.

Chapter 3 will focus on the high-k gate dielectric breakdown under ESD-like stress. This chapter starts with a discussion of the stress configurations on gate dielectric in the context of some real I/O circuits. The characterization results of high-k gate breakdown under similar stress configurations are then presented and compared with SiON gate dielectric from an older technology node. For efficient and accurate characterization, different test methods are discussed and compared as well. Finally, methods to improve the immunity of the I/O circuits to gate dielectric breakdown are proposed.

Chapter 4 reports the degradation of NMOSFETs under different levels of non-destructive ESD-like stresses. The effects of both positive and negative stress applied on the gate are discussed first, which emulate ESD events on the input receiver pin. The effect of the stress on threshold voltage V_t and drain saturation current I_{dsat} and the stress induced damage to the gate dielectric/Si interface are presented. The performance degradation is compared with devices with SiON gate. Moreover, the

impacts of the stress on the high-k gate leakage current and on the positive-bias-temperature instability degradation kinetics [3] are also covered. Finally, the effects of positive stress applied on the drain are discussed, which emulates ESD events on output pins.

Chapter 5 discusses the design and optimization of the SOI field effect diode (FED), using a TCAD simulation based approach. TCAD simulations help identify important device parameters. Based on the simulation results, methods to increase the anode breakdown voltage (V_{FB}) to the level which is compatible with supply voltage in the current technology are proposed. This chapter concludes with a brief discussion on the experimental verification of the proposed methods conducted by another research group.

Finally, Chapter 6 summarizes the contributions of this dissertation and discusses the scope of future work.

[3] Please refer to Sec. 2.4.5 for details.

Chapter 2: Review of ESD and gate dielectric breakdown and degradation

Electrostatic Discharge (ESD) is a single-event, rapid transfer of electrostatic charge between two objects, usually resulting when two objects at different potentials come into direct contact with each other. ESD can also occur when a high electrostatic field develops between two objects in close proximity. A familiar example of electrostatic charging and discharging is when someone walks across a carpet and then touches a metal door knob. The charging level can be as high as 35000 V in a dried environment. The ESD events are usually harmless in everyday life since the total amount of charge is typically small. However, they are one of the most serious reliability concerns in the semiconductor industry. ESD damage due to handling and testing is directly responsible for approximately 10% of the total failure returns [22, 23]. It has negative influence on product yield [23] and becomes more prevalent as the device dimension keeps scaling down. To protect ICs from these undesirable ESD events, ESD protection networks are required at every pin of the chip.

There are several organizations such as electrostatic discharge association (ESDA), electronic industries alliance/joint electron device engineering council (EIA/JEDEC) that issue ESD test standards to predict the ESD immunity level. The most common industrial models used to measure ESD robustness are the human body model (HBM), the machine model (MM) and the charged device model (CDM). This chapter begins with a discussion of these models, followed by a detailed description of the transmission-line pulsing (TLP) characterization method. An overview of the basic

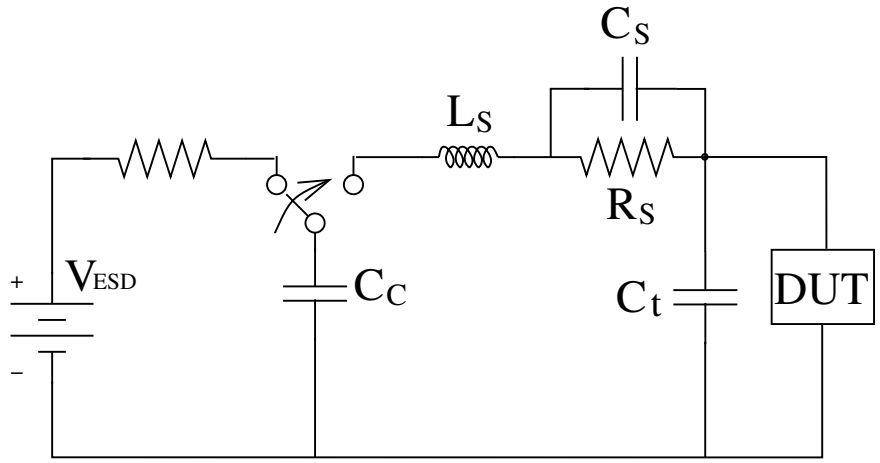


Figure 2.1: The lumped circuit model for the HBM and MM. Capacitor C_c is charged to the test voltage, V_{ESD}, and then discharged through R_s to the device under test (DUT) by closing the switch. Parasitic circuit elements are represented by series inductance L_s, stray capacitance C_s, and test-board capacitance C_t. The typical circuit parameters are (a) $C_C = 100$ pF, $L_s = 7.5$ μH, $R_S = 1500$ Ω and $C_s = 1$ pF for HBM and (b) $C_C = 200$ pF, $L_s \approx 2.5$ μH, $R_S = 5$ Ω and $C_s = 1$ pF [6].

ESD protection strategies will be discussed next. Finally, the physics of the ESD related reliability issues including gate dielectric breakdown and latent damage will be reviewed.

2.1 ESD Characterization Models

2.1.1 Human Body Model

The human body can be charged and then transfer the charge to a semiconductor device through normal handling or assembly operations. To evaluate the effectiveness of the protection circuitry in an integrated circuit, Human Body Model (HBM) ESD testing is performed. This model represents an ESD event from a charged human body to the pin of an integrated circuit.

The circuit used to model the HBM event is shown in Fig. 2.1 [6,24,25]. The pulse is generated by charging the capacitor (C_C) of 100 pF and then discharge it through

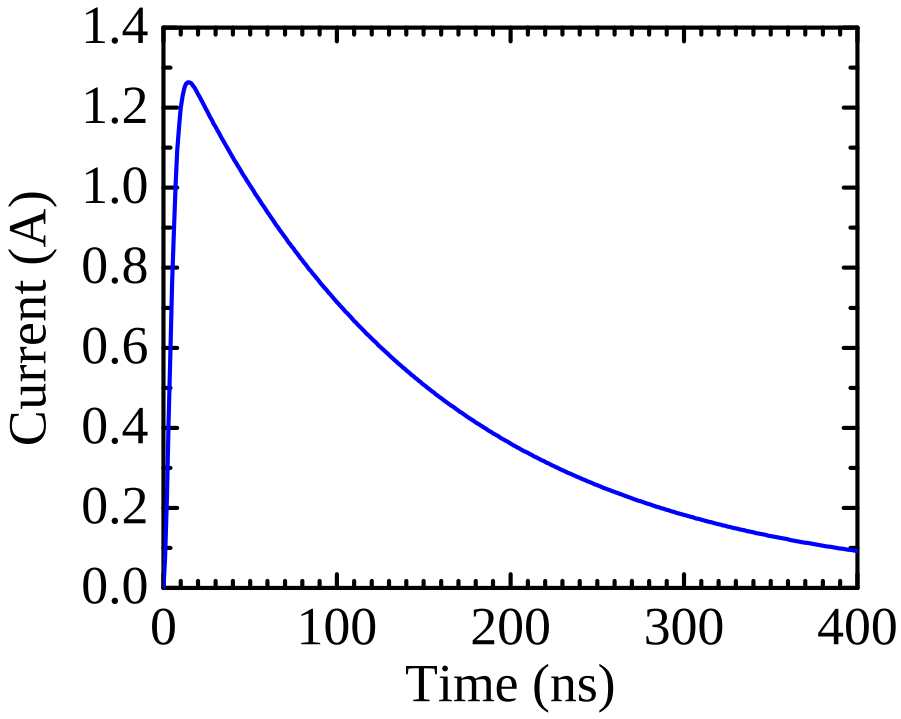


Figure 2.2: A typical waveform for a 2000 V HBM event [6].

a 1.5 kΩ resistor (R_s) into the Device Under Test (DUT). Because of the large series resistance R_s in the HBM, this ESD event can be modeled as a current source with the typical peak value of 1.20-1.48 A for 2 kV HBM ESD stress. The rise time is 2-10 ns and the decay time is 130-170 nm. Testing specifications usually require circuits to be able to have a minimum threshold of +/- 2kV HBM ESD stress on all pins. HBM is the longest ESD event of the three primary models, but it has the lowest current. A typical HBM pulse is shown in Fig. 2.2, obtained from SPICE simulation using the circuit model shown in Fig. 2.1.

2.1.2 Machine Model

The Machine Model (MM) is similar to the human body model. A capacitor is charged to a certain voltage and then discharge through the device. The equivalent lumped

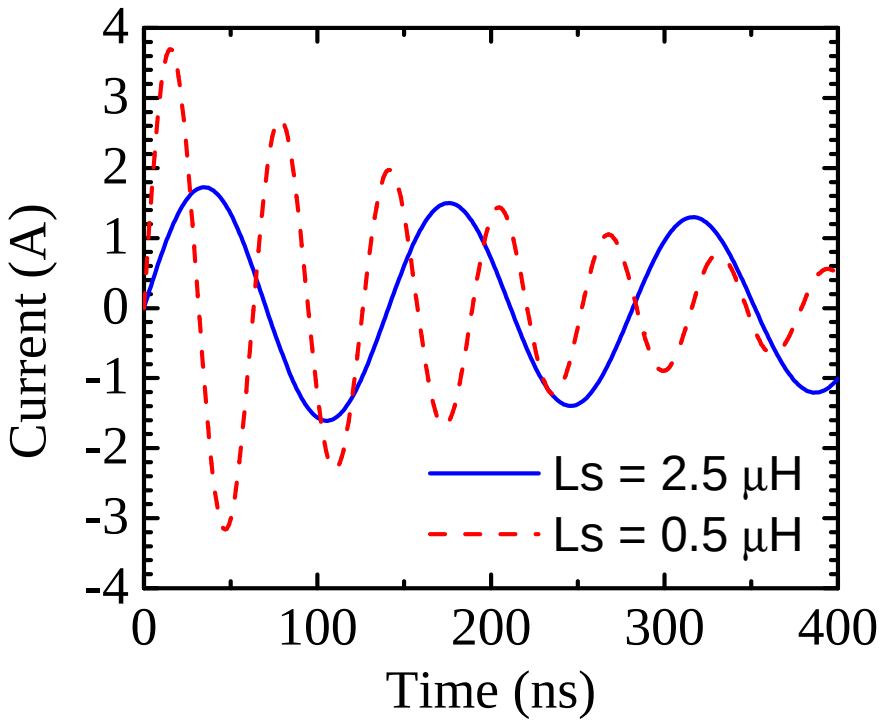


Figure 2.3: A typical waveform for a 200 V MM event with two different series inductance [6].

circuit model is similar to the shown in Fig. 2.1 with modified component parameters. The capacitor C_C is increased to 200 pF, the parasitic resistance R_s is decreased to a few ohms and a series inductance L_s of 1 μH is used. A typical waveform is shown in Fig. 2.3 using the typical parameter values shown in Fig. 2.1. This model is intended to simulate the type of damage caused by equipment used in manufacturing. As shown by the equivalent circuit model, the higher charging capacitance and lower impedance of the path result in higher discharging current densities during the MM discharge an a shorter rise time compared to the HBM event. The very low series resistance also implies that the dynamic impedance of the device under test and the values of the parasitic capacitance and inductance have a large effect on the MM waveform, as illustrated by Fig. 2.3. Changing the series inductance L_s from 0.5

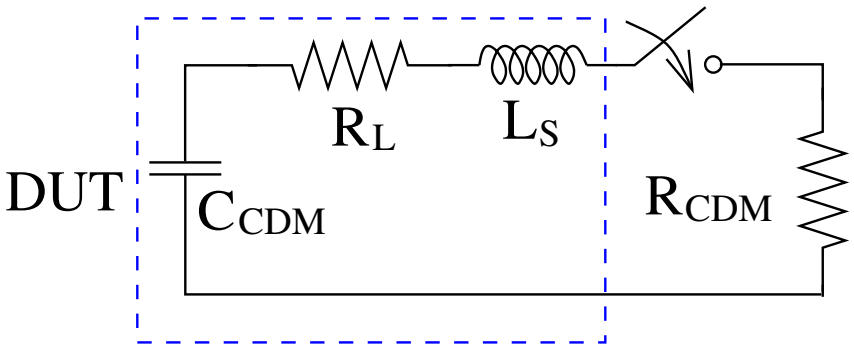


Figure 2.4: The equivalent circuit model for CDM. C_{CDM} is the sum of all capacitances in the DUT and the package with respect to ground and R_{CDM} is the total resistance of the discharge path. C_{CDM}, R_L and L_S represent the DUT [7]. The device under test is charged and then discharged through a pin to ground. For 500 V CDM, the typical model parameters are $C_{CDM} = 10$ pF, $R_{CDM} = 10$ Ω, $R_L = 10$ Ω and $L_s = 10$ nH.

μH to 2.5 μH leads to drastic changes in the rise time and peak current. The pass threshold for the MM required for the commercial ICs is ~ 200 V.

2.1.3 Charged Device Model

The Charged-Device Model (CDM) is intended to simulate the situation when the packaged IC gets charged up and then the stored energy is discharged through a low-inductance, low-impedance path when one of the pins is grounded. The equivalent circuit model is shown in Fig. 2.4 [7]. This model is major concern today because of its very fast rise time which is about 100-200 ps and its large peak current value which is typically several amps. Fig. 2.5 is a typical CDM waveform, obtained using the typical parameter values shown in Fig. 2.4. Since the turn-on time of the protection circuits is on the order of 1 ns, pad voltages may become too high during a CDM event and thus damage the thin oxides of the MOSFETs. Therefore the turn-on speed and the on-resistance of the ESD clamp device are very important for robust CDM ESD protection as the CMOS scaling continues.

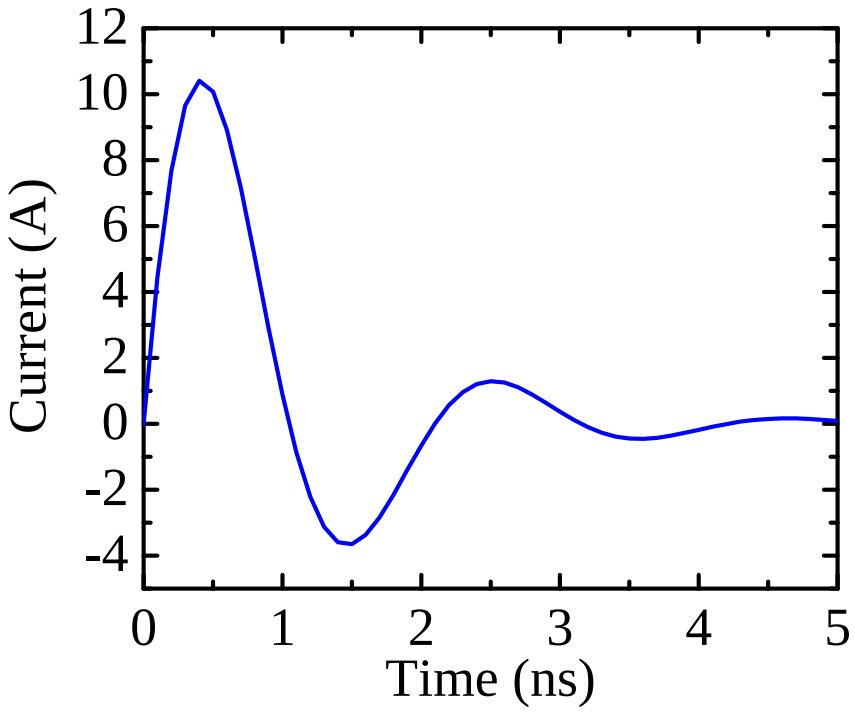


Figure 2.5: A typical waveform for a 500 V CDM event.

2.2 Transmission Line Pulsing method

The classical ESD models (HBM, MM and CDM) have many limitations to characterizing ESD robustness of circuits. The complex pulses of these models make the response of the circuit complex and thus difficult to measure. Therefore these models offer only restricted insight in how the protection circuits work and how and where they fail.

To avoid these difficulties, the Transmission Line Pulsing method (TLP) was introduced by Maloney and Khurana [26] as a new electrical analysis tool to test the many single elements used as ESD protection structures. Instead of duplicating the real life ESD events such as HBM, MM or CDM, a series of square wave pulses of varying magnitude and a constant pulse width, as shown by Fig. 2.6, are applied to

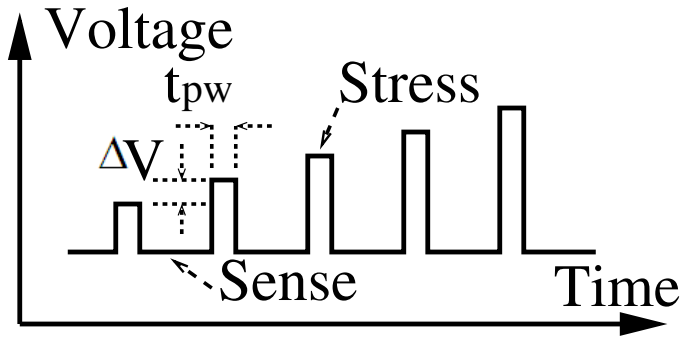


Figure 2.6: TLP uses a series of square wave pulses of varying magnitude and a constant pulse width to stress the device.

stress the DUT and the current through and voltage across the device are measured. As TLP uses a reliable and repeatable sqaure waveform to stress the device, it can be used to obtain current-voltage relationship for the semiconductor devices and circuits under different ESD conditions. After the application of each ESD stress pulse, the DC leakage current of the DUT is recorded. The TLP I-V along with the DC leakage current (or, device I-V) data provide electrical indications of where damage begins, and how rapidly it can evolve from soft to hard failure. The rectangular TLP pulse simulates the energy of different ESD events. Different pulse widths can be used to simulate different events. Typically, a 100 ns pulse corresponds to HBM, and a 30 ns pulse is used to simulate MM and 1 ns to simulate CDM.

2.3 ESD protection strategy

ESD event can occur between any two pads/pins of an IC. Therefore ESD protection devices are required for all pads/pins. These devices remain off during the normal operation conditions and have negligible effects on the circuit operation. During ESD events, these devices turn on and form a low resistive path. As a result, the ESD current is shunt to ground through these protection devices and hence reduces the flow of ESD current through the driver or receiver circuit components. Most ESD

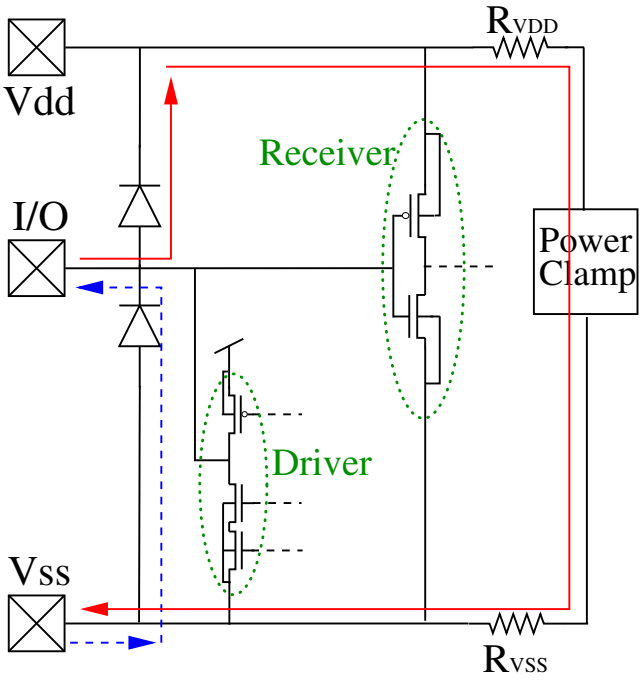


Figure 2.7: An input/output circuit based on double-diode ESD protection strategy. Solid lines with arrows indicate the designed ESD discharge path under positive ESD events, while dashed lines represent the designed discharge path under negative ESD events. The arrows indicate the direction of the current flow. R_{VSS} and R_{VDD} are the resistance of the supply bus and ground bus respectively.

solutions rely on shunting current from an I/O pin to a power supply rail and then the charge can be distributed to other I/O pins or supplies [27]. These solutions fall into two categories: (a) Vss-based ESD protection that shunt current to the negative supply rail and (b) Vdd-based ESD protection that shunt current to the positive supply rail. As an example, the ESD related charges are assumed to be shunted to the negative supply rail (the Vss rail) in the following discussions.

2.3.1 Double diode based ESD protection strategy

One of the most commonly used ESD protection strategies is to connect two diodes to the pin, one to the positive supply bus and the other to the negative supply bus, as shown by Fig. 2.7. The current paths in the chip for the possible discharges

between the pin and its negative supply bus are also shown in this figure. The solid lines with arrows represents the designed discharge paths under positive ESD events. For positive pulse with respect to Vss, the current passes through the upper diode, along the positive supply rail to the power clamp, through the clamp and to the Vss supply rail. The dashed lines with arrows indicate the designed discharge paths under negative ESD events. For a negative pulse with respect to Vss, the current passes through the lower diode and out the signal pin.

Good ESD results are achieved when the discharge paths are confined to these routes. If the resistance of the wiring, the power clamp or the protection devices (i.e. the diodes here) causes the signal pad or Vdd bus to rise too high in voltage, the discharge path may be rerouted through I/O or internal circuitry. Therefore the discharge path must not only be able to handle the current, it must be also able to do so without allowing large voltages to develop across the parallel circuitry due to I-R effects. Because the diodes appear as parasitic capacitance on the I/O node, the designer may have to balance ESD protection against circuit performance. The above two points are the general rules which should be followed in the design of all ESD protection networks.

2.3.2 Local clamp device based ESD protection strategy

Another typical ESD protection circuit topology is based on a local clamp device, as shown in Fig. 2.8. The local clamp devices are typically snap-back devices, including non-silicided grounded gate NMOSFET, silicon controlled rectifier (SCR), etc. They are connected between the I/O pad and the Vss rail. For a positive pulse with respect to Vss, the local clamp device turns on and shunts the current directly to the Vss rail. For a negative pulse on the I/O pad with respect to Vss, the current passes through the ESD diode in parallel with the local clamp device and then out

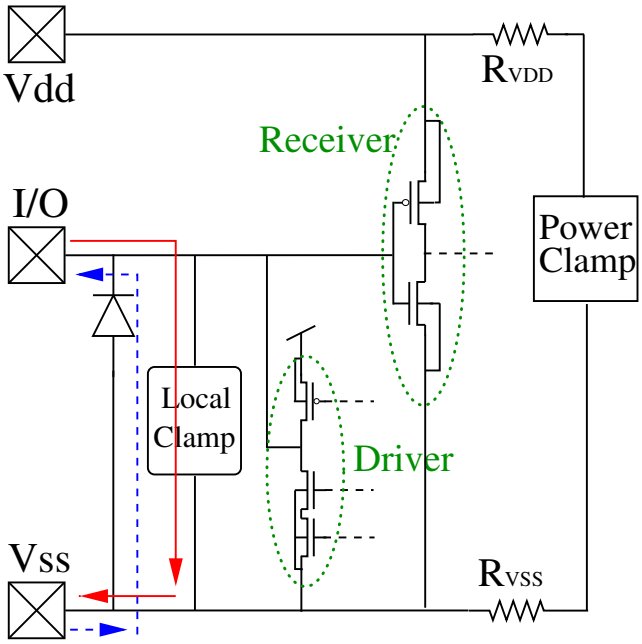


Figure 2.8: Local clamp device based ESD protectoin strategy.

the I/O pad. An advantage of this topology is that the I/O pad can rise above the power supply voltage without forward-biasing the upper diode, which is useful for mixed voltage interface applications or hot-pluggable interface applications. Another important advantage of the local clamp based ESD protection will be discussed in the Sec. 2.3.4. However, when implementing this protection strategy, it is necessary to ensure that the protection device's turn-on voltage is higher than Vdd so as to avoid accidental turn-on and reduce leakage current during normal operation.

2.3.3 CDM protection strategy

During a CDM event, the entire chip is charged and then discharged through a single pin. For the circuit shown in Fig. 2.7, assume the chip is charged to a negative voltage and then discharges through the I/O pin and all the buses and devices are at the same potential (the circuit is unpowered). The preferred charge flow path is through the upper diode, along the Vdd bus to the power clamp, through the clamp to the Vss

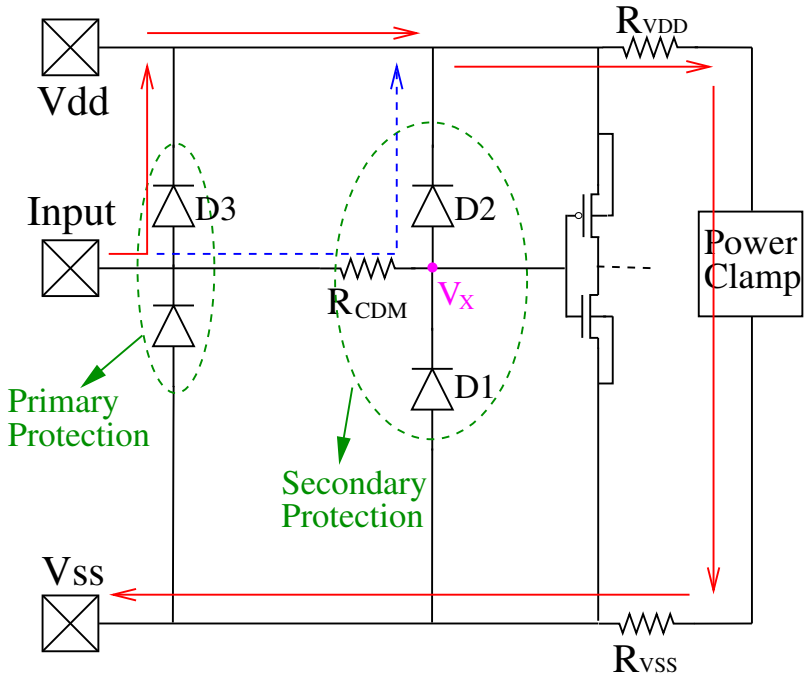


Figure 2.9: Basic CDM protection strategy

rail. For a very robust ESD protection diode, the on-resistance is about 1 Ω. The CDM current can be in the range of 10 A. This leads to a 10 V drop across the upper diode, also across the gate dielectric the MOSFETs in the receiver. This voltage is large enough to break down the gate oxide and destroy the receiver.

To improve the CDM robustness, a two-stage protection scheme, a primary protection stage along with a secondary protection stage as shown in Fig. 2.9, is normally used. Consider the same condition as described in the previous paragraph. The majority of the discharge current will flow through the diode D3, while a small portion will flow through the CDM resistor R_{CDM} and the diode D2. Although the voltage drop across the diode D3 is still large ($\sim$ 10 V), the voltage (V_x) at the receiver MOSFET gate is small because of the voltage divider formed by R_{CDM} and D2. In this way, the gate oxide is protected against CDM events.

2.3.4 Necessity of Local Clamping

The double-diode based clamping scheme is becoming insufficient to protect the gate dielectric especially during the high-current CDM event. The ESD current path in this case is very long, involving the Vss and Vdd bus resistance, the power-clamp, and the ESD diodes. The excess voltage buildup along this long path may overstress the transistor gate dielectric and induce damage or failure. To improve the protection effectiveness against the high-current CDM events demands the exploration of new design strategies and devices to avoid high voltage at the I/O pad. The local clamping protection strategy shown in Fig. 2.8 is more appropriate for CDM protection compared to Fig. 2.7. By connecting a local clamp device directly between the pad and Vss, the protection circuit allows ESD current to flow from the pad to Vss rail without going through the Vdd power buses and the power clamp between the positive and negative supply rails. Consequently, the discharge path resistance is appreciably reduced and the pad voltage build-up is minimized.

2.4 Background on gate dielectric degradation and breakdown

Catastrophic failures under ESD stress include gate dielectric failure, metal melting and junction filamentation. In these cases, the device is permanently damaged due to the high power dissipation. In contrast to catastrophic failure, ESD stress may also degrade device performance and thus degrade the circuit performance. We will discuss the gate dielectric breakdown and device degradation of MOSFETs with high-k gate under ESD like stress in Chapter 3 and Chapter 4. In this section, a brief introduction to the gate dielectric breakdown and degradation will be presented.

2.4.1 Necessity of introducing high-k gate dielectrics

Because of the excellent quality of SiO_2 grown on a Si substrate (SiO_2/Si) with optimized fabrication processes, SiO_2 has been used predominantly for the gate insulator of MOSFETs over the last 40 years. The large barrier height ($\sim$ 3 eV) of the SiO_2/Si interface leads to much reduced leakage current for sufficiently thick films. The high melting temperature of SiO_2 makes it very compatible with CMOS processing steps performed after the formation of the SiO_2-based gate dielectrics.

The MOSFET on-current can be expressed as

$$I_{ON} = \frac{\mu_{eff} C_{ox}}{2} \frac{W}{L} (V_{GS} - V_t)^2, \tag{2.1}$$

where μ_{eff} is the effective channel mobility, W is the width of the device, L is the lenght of the channel, V_{GS} is the gate voltage with respect to source and V_t is the threshold voltage of the MOSFET. C_{ox} is the capacitance per unit area of the gate dielectric expressed as

$$C_{ox} = \frac{\kappa \epsilon_0}{t_{ox}}. \tag{2.2}$$

ϵ_0 is the vacuum permittivity and κ is the relative permittivity of the gate dielectric. Based on the constant voltage scaling rule shown in Table 1.1, for every successive technology node, the device dimensions (W, L, t_{ox}) are scaled by a factor of k; the voltage is also scaled by the same factor. Eq. (2.1) and Eq. (2.2) indicate that the intrinsic delay time CV/I_{ON} is decreased by a factor of k with technology scaling. Therefore for better performance, t_{ox} has been scaled down for many years. However, the reduced thickness of SiO_2 results in excessively large gate leakage current originating from the quantum mechanical tunneling effects. Further scaling the oxide thickness will make the gate leakage current intolerably large, as shown in Fig. 2.10

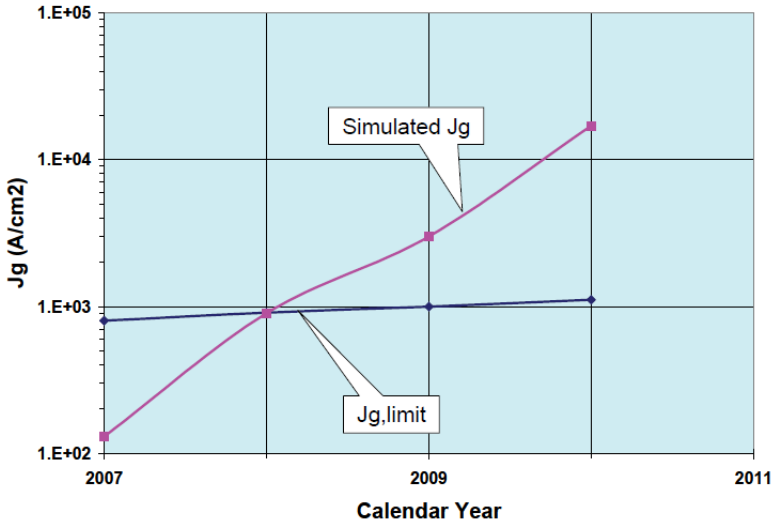


Figure 2.10: High-Performance Logic: gate current density limit versus simulated gate leakage current density for SiON gate dielectric [8].

predicted by [8]. This large gate leakage current stops further scaling of t_{ox}.

Eq. 2.2 also indicates that the targeted C_{ox} can be also achieved by increasing the relative permittivity of the dielectric material κ while further reducing the thickness t_{ox} . Consequently, to keep improving circuit performance, there has been substantial efforts to develop high permittivity, or high-k, dielectric materials to replace SiO_2. Table 2.1 compared the potential high-k materials and their relative permittivities to SiO_2. The hafnium-oxide-based materials (HfO_2, $HfSi_xO_y$, HfO_xN_y, $HfSi_xO_yN_z$) have already been used in the 32 nm technology node to replace SiO_2. However, none of these materials have been found to possess the quality as good as SiO_2. To improve the Si/gate dielectric interface, a thin SiO_2 layer is typically grown as the buffer layer (or interfacial layer-IL) before the deposition of the high-k material to form the final gate stack.

Table 2.1: Dielectric materials and their relative permittivities [3].

Material	Relative permittivity κ
SiO_2	3.9
Al_2O_3	9
HfO_2	25
Ta_2O_5	26
ZrO_2	29
La_2O_3	30
Pr_2O_3	31
TiO_2	95
$SrTiO_3$	200

2.4.2 Physical models for gate dielectric degradation and breakdown

Under electrical stress for a certain period of time, gate dielectric will wear out and then breakdown. In other words, the dielectric film loses its insulating properties and becomes conductive. This phenomenon was first observed more than three decades ago [28, 29]. The gate conduction can cause malfunction of circuits employing field effect transistors. It is currently widely accepted that oxide breakdown is related to defects generated in the oxide during electrical stress.

A variety of phenomena related to the oxide degradation have been observed, such as positive charge trapping; generation of neutral electron traps (and the related trapping of electrons); generation of Si/gate dielectric interface states and increase of the low-field leakage current. All of these effects have been used to monitor the degradation of gate dielectrics under different stress conditions. However, the exact physical mechanisms that determine the degradation and the eventual breakdown are still unclear. There are still many open questions about the actual physics involved. Several different physical models have been proposed to explain defect generation, for example:

- The Thermochemical Model [3, 30–34];
- Anode Hole Injection model (AHI) [35–39];
- Anode Hydrogen Release model (AHR) [40–42].

For the thermochemical model, the defect generation is attributed to the high electric field across the gate dielectric. The electric field interacts with the local dipole moment and reduces the activation for chemical bond breakage. This leads to enhanced defect generation during stress. However, this physical picture fails to explain many recent experimental results. For example, the thermochemical shows that the breakdown is a field- and time-driven process while the contribution from the tunneling current and the carrier energy is not important. However, compelling experimental evdicences [38, 43, 44] indicates that both electron fluence and energy play an important role in the breakdown process. These observations have led to the other two models based on the injection of carriers through the oxide and on the dissipation of electron energy near the anode interface. The hot electrons release some specie (hydrogen assumed in AHR and hole assumed in AHI) near the anode. Part of the released species diffuse or drift back through the gate dielectric and react with defect precursors to form oxide defects.

2.4.3 Gate oxide breakdown statistics

As the breakdown is due to random defect generation, the breakdown process is a stochastic process, rather than a deterministic phenomenon, which means that the values of breakdown field, breakdown voltage, or time-to-breakdown will be different if we repeat the same stress experiment in nominally identical samples. In the literature, it is generally accepted that the dielectric breakdown process follows the

two-parameter Weibull distribution,

$$F(t_{BD}) = 1 - \exp\left(-\frac{t_{BD}}{t_{63\%}}\right)^{\beta}, \tag{2.3}$$

where F is the cumulative failure probability and t_{BD} is the time-to-breakdown. The distribution has two important parameters: β is the shape factor and $t_{63\%}$ is the characteristic time-to-breakdown at the failure percentile of 63.2%. Rearranging Eq. (2.3) into the following form:

$$W = \ln\left[-\ln(1 - F(t_{BD}))\right] = \beta \times (\ln t_{BD} - \ln t_{63\%}), \tag{2.4}$$

and plotting W versus $\ln(t_{BD})$ yield a straight line with slope β, as shown in Fig 2.11(a). To assure the reliability of the product, accurate data at the low failure percentile, for example $F = 0.01\%$, is of crucial importance. However, it is impractical to measure a large amount samples to approach the low failure percentile range. It is crucial to predict $t_{63\%}$ and β precisely based on a limited sample size. Various methods have been proposed for this purpose [9, 45]. Reducing the gate dielectric thickness not only decreases the characteristic breakdown time but also reduces the slope β as shown in Fig. 2.12. On the one hand, smaller β implies the breakdown data becomes more widely spread, which makes characterization more difficult. On the other hand, smaller β means shorter lifetime for the same $t_{63\%}$. Therefore, the reliability margin of the gate dielectric significantly reduces with scaling.

Oxide failure probability can also be described well by randomly distributed defects following the Poisson model [9]

$$F = 1 - \exp(-DA), \tag{2.5}$$

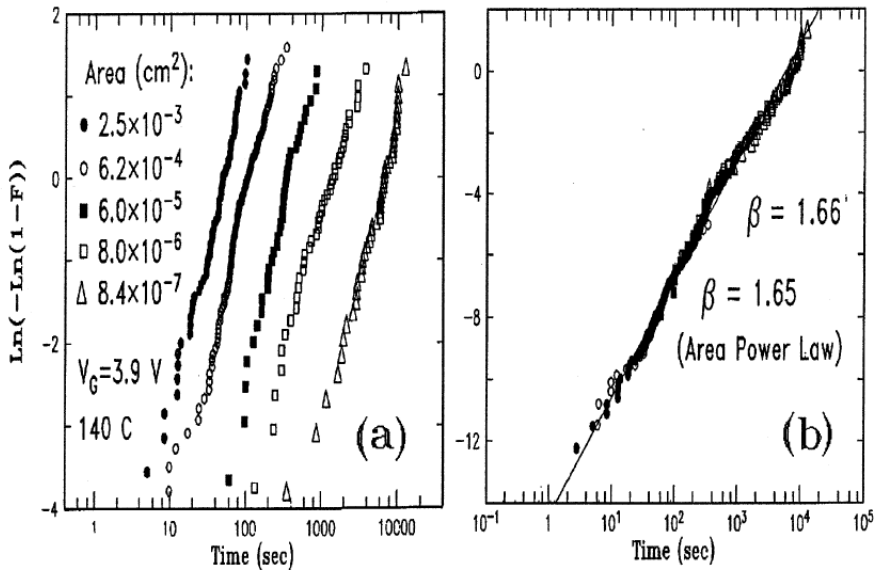


Figure 2.11: Demonstration of Weibull plot and the area scaling law [9]. (a) t_{BD} distributions of SiO_2 gate dielectric with the same insulator layer thickness (2.66 nm) but different gate dielectric areas. (b)t_{BD} distributions after being scaled using Eq. (2.6) to the reference area of 8.4e-7 cm^2.

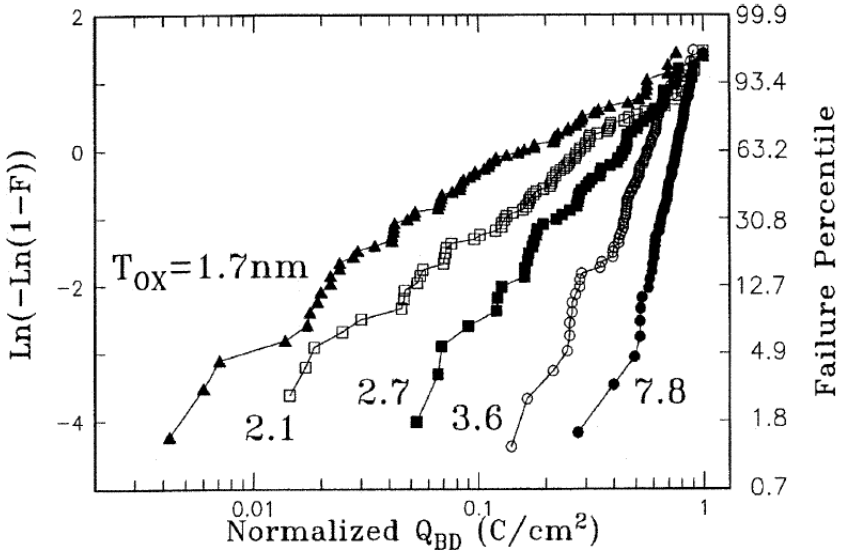


Figure 2.12: Weibull shape factor (Weibull slope) increases with oxide thickness. [9].

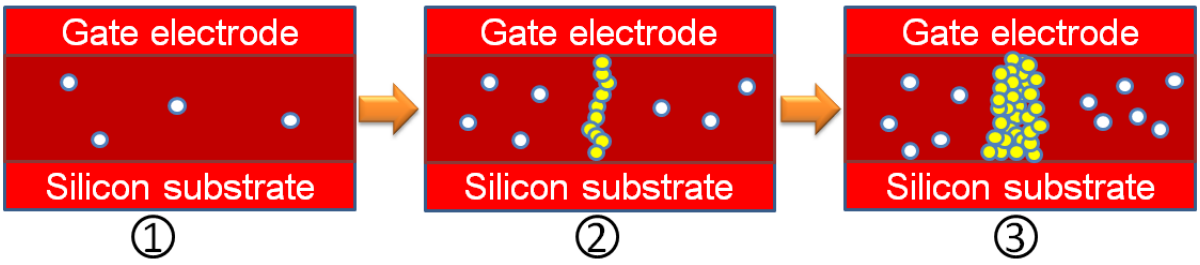


Figure 2.13: Oxide breakdown can be divided into 3 stages: (1) Defect creation; (2) Percolation path created; (3) Wearout of the percolation paths.

where A is the gate oxide area and D is the defect density. From Eq. (2.5) and Eq. (2.3) can be seen that the characteristic breakdown times $t_{63\%,1}$ and $t_{63\%,2}$ of two distributions follow the area scaling law

$$\ln\left[-\ln\left(1-F_1\right)\right]-\ln\left[-\ln\left(1-F_2\right)\right]=\ln(A_1/A_2). \tag{2.6}$$

An example of area scaling is shown in Fig. 2.11. As the reliability measurement is usually performed on individual transistors, this law is very useful for the lifetime prediction of a whole chip of which the gate dielectric under stress corresponds to the gate area of thousands of transistors.

The origin of the breakdown statistics described above can be explained by the percolation theory [46–50], which states that the breakdown occurs when the created defects form a conduction path (a percolation path) linking the gate electrode to the silicon substrate. The properties of the the breakdown statistics, such as the Weibull distribution, area scaling and oxide thickness dependent Weibull shape factor β, as predicted by the percolation model agrees very well with the experimental results.

2.4.4 Post breakdown mode

The breakdown process described by the percolation theory can be divided into three stages as shown in Fig. 2.13:

1. Defect creation: as the stress starts, defects are created at random locations inside the gate oxide. These defects increase the gate leakage current slightly (Stress Induced Leakage Current, SILC) as shown in Fig. 2.14.

2. Breakdown path creation: breakdown occurs when the created defects form a conduction path linking the Si substrate to the gate electrode.

3. Breakdown path wearout: the created breakdown path wears out, leading to an abrupt increase in the gate leakage current or a progressively increased gate leakage current.

How fast the percolation path wears out and the time dependence of the post-breakdown gate leakage current depend on a variety of factors. For example, it depends on the thickness of the gate dielectric. For thick SiO_2 (> 5 nm), the breakdown is usually abrupt and the transition from stage 2 to stage 3 is not observable typically. For thin SiO_2 gate dielectric (in the range from 3 to 5 nm) with polycrystalline silicon gate, after the formation of the percolation path, the gate leakage current increases slightly and the I-V characteristics can be fitted to a power law or an exponential law. This is called Soft Breakdown (SBD), because this will not necessarily lead to malfunction of the circuit. Then I_G stays more or less constant with increased noise until a large abrupt increase in I_G occurs, after which the I-V characteristics can be fitted to a linear relationship. This event is called Hard Breakdown (HBD) since the magnitude of I_G is large enough to cause circuit malfunction. An example of SBD and HBD is shown in Fig. 2.14, where SBD and HBD can be clearly

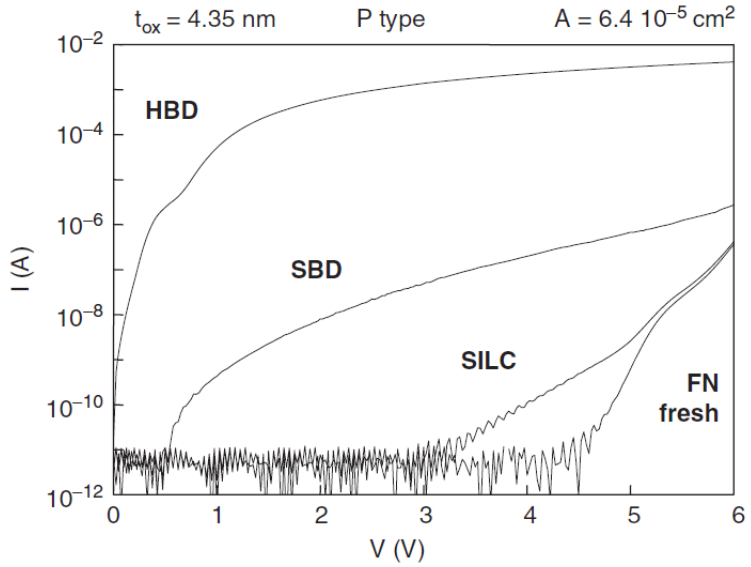


Figure 2.14: IV characteristics of gate leakage current at different stages of oxide degradation. Fresh IV, stress induced leakage current (SILC), and typical post-SBD and post-HBD IV curves are shown [10].

identified. For ultrathin gate oxide (thinner than 2 nm), SBD and HBD becomes very hard to distinguish, because the SBD is not stable and the HBD is progressive. Moreover, the background tunneling current becomes so large that it may mask the SBD event. Fig. 2.15 gives some examples of the evolution of the post BD gate currents.

2.4.5 TDDB and latent oxide damage under ESD-like stress

TDDB under ESD-like stress

Three different stress methodologies are generally used to study Time-Dependent Dielectric Breakdown (TDDB) under ESD pulses [51, 52].

1. Constant voltage stress: a long pulse is applied to the gate and the gate oxide fails within the duration of only a single pulse. This method gives directly the time-to-breakdown of the gate dielectric under a specific stress voltage. In order to get the breakdown voltage corresponding for different ESD events (which

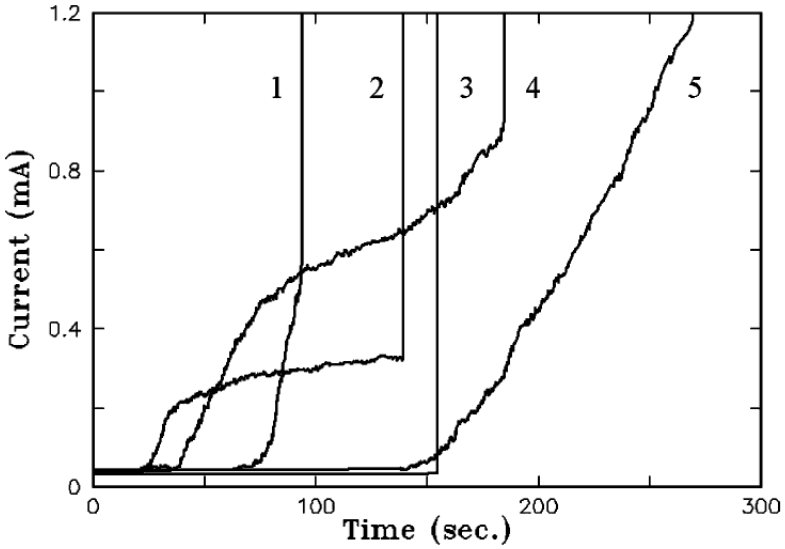


Figure 2.15: Evolution of the post-BD current in several 8 mm^2 NFETs with a 1.35 nm-thick oxide stressed under CVS at $V_G = 2.75$ V [11].

corresponds to different pulse duration), time-to-breakdown at several different stress voltages needs to be characterized so as to establish a mathematical relationship between time-to-breakdown and the stress voltage. In this way, the breakdown voltage for different ESD events can be calculated according to their different time durations (HBM-100ns, MM-30nm and CDM-1ns).

2. Repetitive constant voltage stress (RCVS): a sequence of short pulses is applied to stress the gate oxide. Oxide does not breakdown within the duration of one single pulse. DC leakage or other measurements may be done between each pulse. Time-to-breakdown is directly characterized and is defined as the number of pulses applied multiplied by the pulse width plus the the time to breakdown in the final pulse;

3. TLP method or Voltage ramped stress (VRS): devices are stressed by a sequence of pulses with the same width but increasing magnitude until the oxide breakdown. This is similar the TLP method used in the standard ESD tests

mentioned above. t_{BD} can be derived following the procedure described in [53].

So far, very few studies have been done to explore TDDB under ESD-like pulse stress and the study only focused on SiO_2 or SiON gate dielectric. Cheung's results [54] indicates that gate oxide degradation mechanism remains unchanged in the short time scale, consistent to the study on oxide breakdown transients by Lombardo [55]. Then Matsuzawa *et al.* [56] studied a 3 nm oxide using the AHI model and showed that the breakdown remains unchanged over the range of 10^{-4} to 10^{-7} s. To understand TDDB of thin-oxide under ESD like stress, Wu *et al.* [57, 58] studied PMOS capacitors (2.2-4.7 nm) in accumulation ($+V_G$) over the range of $10^2 - 10^{-8}$ s. They found that time-to-breakdown has a $1/E$ dependence due to Fowler-Nordheim (F-N) dominated tunneling current during the short stress. They also pointed out that self-heating plays a negligible role in the TDDB measurement due to the small time scale of the ESD events. Later on, Weir *et al.* [51,59] reported results on NMOSFET and PMOSFET in inversion with 1.5 nm SiON gate insulator using TLP method. Based on their results, $t_{63\%}$ in the ESD time domain can be extrapolated from DC measurement results using the voltage power law [60]. More recently, Ille *et al.* [52,61] also concluded that for the oxide (1-7 nm), the voltage power law holds over the range 10^4 to 10^{-8} s. Additionally, the authors also compared the three stress methodologies (CVS, RCVS and TLP) and claimed that CVS should be used as the characterization technique. Chen *et al.* pointed out that gate dielectric failure sometimes does not cause circuit function failure [62]. They noticed that for transistors from 90 nm and 130 nm technology nodes, soft breakdown and hard breakdown can be clearly identified in the sub-ns time domain [12]. Their results as in Fig. 2.16 shows that the gate leakage resistance has three different plateaus: 10 GΩ for fresh devices, 1 MΩ for post-SBD and 3 kΩ for post-HBD. In their later work [63,64], the authors developed a

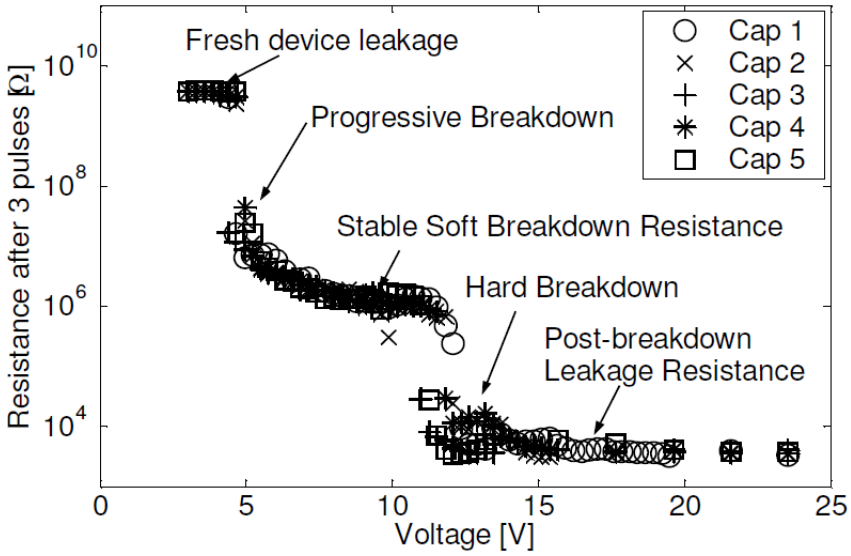


Figure 2.16: Gate leakage resistance after the 175 ps pulse for five different instances of the same capacitor under different stress voltages [12].

marco model for post-breakdown transistors, which makes fast post-breakdown simulation possible and helps avoid over-design. In summary, it is accepted that the oxide breakdown event could be completed within the time scale of an ESD event; and the statistics of oxide breakdown is retained as well. However, the relation between $t_{63\%}$ and stress voltage, and the influence of different stress methodologies are still unclear. Besides, the introduction of high-k/metal gate stacks opens more questions, on which no work has been done yet to the best of our knowledge.

Latent damage caused by ESD-like stress

As mentioned in the previous section, defects are created well before oxide breakdown. These defects degrade the device performance and may affect the long-term reliability of transistors. Consequently, another topic drawing much attention is to study the latent damage caused by the ESD-like stress.

Aur *et al.* [65] have shown that ESD pre-stressing induced latent damage could decrease the hot carrier injection (HCI) lifetime by as much as a factor of 4. It

has also been reported by Doyle *et al.* [66] that NMOSFETs show series resistance like changes due to the injection of charge carriers in the oxide during high current stress. Recently, Salman *et al.* showed that ESD pulses on the drain cause a certain amount of Hot Electron (HE)-like degradation [67] and also degrade the oxide in terms of stress induced leakage current (SILC) and gate-induced drain leakage (GIDL) [68]. Wu and Rosenbaum [57, 58] studied SILC and demonstrated that the defect generation rate may depend on the pulse width used in the stress. More recently, Cester *et al.* [69] studied latent damages in ultra thin oxide (< 3 nm) by applying 100 ns TLP pulses to either the gate oxide or the drain terminal. It was found that non-destructive TLP pulses on the gate do affect the degradation kinetics of transconductance and the drain saturation current in the subsequent DC stress and the effect is more severe on devices with thinner gate oxide. But the effect on the DC TDDB following the TLP stress is negligible unless the TLP stress voltage is very close to the breakdown voltage (90%). Similar studies was also made by Ille *et al.* [61]. However, it was pointed out in this study that the MOSFET parameters such as saturation threshold voltage is more immune to ESD pre-stress in thin-oxide devices. Besides studying different degradation phenomena under TLP stresses, some researchers also noticed that the type of oxide trapped charges depends on the stress type. Huh *et al.* found that the polarity of the oxide trapped charge may depend on the type of ESD stresses; HBM and MM stresses produce positive trapped charge while CDM stress generates negative trapped charge. Tseng *et al.* [70, 71] compared the trapped charge generated by TLP stress and DC stress. They found that for thin oxide (3.2 nm) the trapped charge are similar between DC and TLP stress while for thick oxide (14 nm) TLP stress generates far less amount of negative oxide trapped charges. All of the above studies are conducted using transistors with SiO_2 or SiON gate dielectric. The degradation of transistors with high-k gate is still unclear.

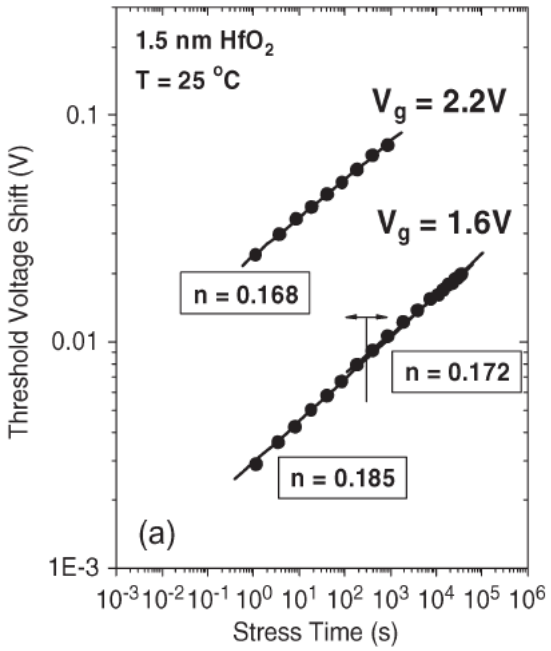


Figure 2.17: Threshold voltage shift during PBTI stress at room temperature for a 1.5 nm thick HfO_2 layer [13].

Moreover, high-k gate dielectric has a specific reliability concern, positive bias temperature instability (PBTI). Even under low positive stress, significant electron trapping is observed, which leads to positive threshold voltage shift, as shown in Fig. 2.17. This degradation phenomena may be related to the presence of a large amount of electron traps, presumably oxygen vacancies [72], in the high-k layer. Therefore, it is necessary to study the effect of ESD-like stress on the performance and reliability of MOSFETs with high-k gate dielectric, especially for NMOSFET under postive ESD-like stress.

2.5 Summary

This chapter presented the preliminary information, including ESD and gate dielectric degradation and breakdown, for a better understanding of this dissertation. The

chapter presents a brief introduction on the ESD phenomena, followed by the description of various ESD models including HBM, MM and CDM, the TLP method as an effective characterization method for different ESD events. Different ESD protection strategies are described along with the requirements for ESD protection networks. This chapter also points out the necessity of introducing high-k gate dielectric. The basic physics of gate dielectric degradation and breakdown are covered. Finally, this chapter reviews the literature results about gate dielectric breakdown and degradation under ESD-like stress.

Chapter 3: Characterization of High-k/Metal Gate Stack Breakdown in the Time Scale of ESD Events

3.1 Introduction

MOSFET gate oxide (GOX) thickness has been reduced for many years to improve device performance. This draws concern on the impact of electrostatic discharge (ESD) events on gate oxide integrity and long-term reliability. Especially for devices in the input/output circuits, the transistors may be connected to external pins, exposing the thin oxide directly to the high-current ESD events. Quite a few studies have been made on the breakdown characteristics of conventional SiO_2/SiON gate dielectric under ESD stress [51,52,58,61,70,71]. However, the excessive off-state gate leakage current caused by direct tunneling through the gate dielectric has stopped the dielectric thickness scaling in MOSFETs. Substantial efforts have been put into the development of high-k dielectric to replace the conventional SiO_2/SiON gate. This opens new questions about gate oxide integrity and reliability. Although there were numerous reports on the breakdown of high-k gate under low electric field stress related to normal device operation[13,73–75], the knowledge about the effect of high-field stress in the ESD time domain is still limited [76], [77].

In this chapter, we investigated high-k gate dielectric breakdown under ESD-like stress. The stress configurations of gate dielectric in the context of real input/output circuits are discussed first. The stress experiments are designed taking into account these different stress condition. We first studied the stress condition for transistors in

input receiver in Sec.3.4. By using different test methods, it is confirmed that high-k dielectric breakdown is catastrophic under ESD-like stress. It is pointed out that the excessive gate current after oxide failure may result in a loss of gate contact and short the drain/source the transistor. Next, the effect of stress interruption on the gate dielectric breakdown is discussed. Using constant voltage stress (CVS) method, the gate oxide breakdown voltages (V_{BD}) of NMSOFETs and PMOSFETs are compared under different stress polarities to identify the worst case scenario. Finally in this section, the results are also compared with SiON gate dielectric. Sec.3.5 discussed the high-k gate breakdown in the scenario of output driver. The results imply that input receiver is more susceptible failure due to ESD induced gate dielectric breakdown. In Sec.3.6, V_{BD} obtained from CVS and from the Transmission Line Pulsing method (TLP) are compared. Finally, Sec.3.7 proposes a method to improve the immunity to gate dielectric breakdown induced failure with the support of experimental results.

3.2 Typical stress conditions on gate dielectric during ESD Events

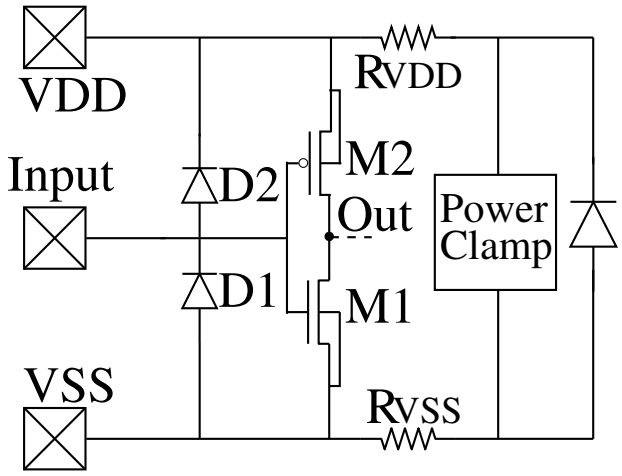
In this section, the input receiver and output driver circuits using either dual-diode or a local clamp device for ESD protection are analyzed under different ESD events. The analysis in this section summarizes different gate dielectric stress conditions and provides proper guidelines for the stress experiments discussed in both Chapter 3 and Chapter 4.

3.2.1 Input receiver under ESD stress

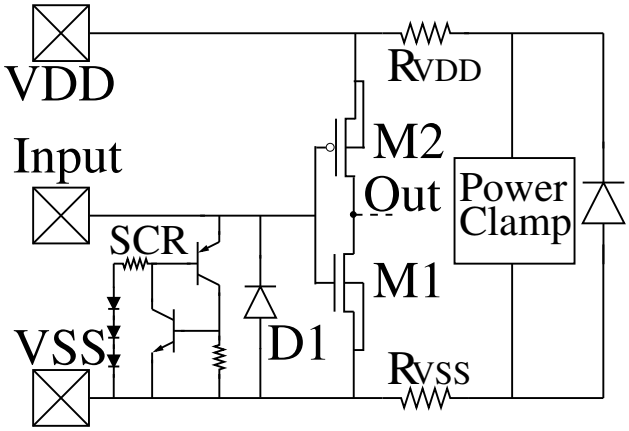
Dual-diode based ESD protection

An input receiver using a dual-diode based ESD protection scheme is shown in Fig. 3.1a. The four ESD discharge modes from the Input pad: positive discharge to VDD (PVDD), negative discharge to VDD (NVDD), positive discharge to VSS (PVSS) and negative discharge to VSS (NVSS), are discussed.

For PVSS, the primary discharge path is through the diode D2, the VDD bus and the supply clamp back to VSS. Because of the voltage drop on D2, the VDD bus resistance R_{VDD} and the supply clamp, the GOX of the NMOSFET M1 and the PMOSFET M2 may be overstressed. Due to the voltage drop on R_{VDD} and the supply clamp, the voltage at the source and body terminals of the PMOSFET M2 is coupled high. Therefore the stress over the GOX of M2 is less severe than that over the GOX of the NMOSFET M1, which is positively stressed with source/substrate grounded and the drain is effectively floating (since the PMOSFET M2 is in the off-state during the PVSS event). NVSS is typically not a concern since the resistance of the primary discharge path (through the diode D1 to VSS) is small and hence D1 is able to clamp the pad voltage low enough. For NVDD, the primary discharge path is through the diode D1, the VSS bus and the supply clamp back to VDD. The negative stress on the Input pad may overstress the GOX of M1 and M2. Because of the voltage drop on the bus resistance R_{VSS} and the supply clamp, the voltage at the source and body terminals of the NMOSFET M1 is coupled low. Therefore the stress over the GOX of M1 is less severe than that over the GOX of M2, which is negatively stressed with the source/body grounded and the drain is effectively floating. Similar to the case of NVSS, PVDD is typically not a concern since the diode D2 is able to clamp the pad voltage low enough.



(a) Dual-diode protection



(b) SCR based local protection

Figure 3.1: Simple receiver circuit with (a) dual-diode and (b) SCR ESD protection.

ESD protection with local clamp device

ESD protection with an SCR-based local clamp device is shown in Fig. 3.1b. The stress configurations during NVSS and NVDD are similar to those with the dual-diode based protection since the primary discharge path is also through the diode D1.

For PVSS, the GOX of the NMOSFET is positively stressed with source/substrate grounded and the drain floating (since the PMOSFET is in the off-state) prior to the turn-on of the SCR. For PVDD, the voltage drop across RVSS and the supply diode

raises the potential at the source and the body terminals of the NMOSFET M1. Therefore, the stress over the GOX of M1 is less severe than that on the GOX of M2 which is stressed by a positive pulse on the gate terminal with source/body grounded and drain coupled to some positive voltage level since the NMOSFET M1 is in the on-state during PVDD.

In summary, the maximum allowable pad voltage is limited by the following stress conditions: (a) the NMOSFET and PMOSFET may be stressed in inversion with the source/substrate grounded and the drain floating; (b) the PMOSFET gate may be under positive stress with source/substrate grounded and drain coupled to some positive voltage level.

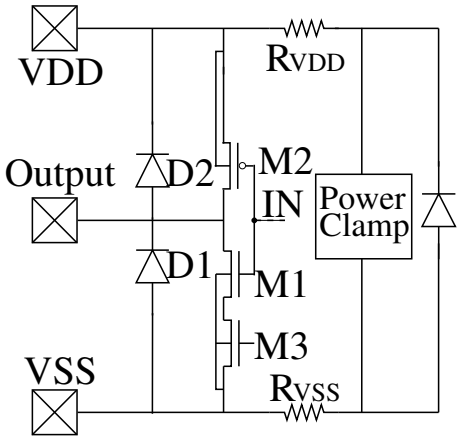
3.2.2 Output Driver under ESD Stress

The output drivers with dual-diode and local ESD protections are shown in Fig. 3.2a and Fig. 3.2b, respectively. To ensure that the parasitic bipolar transistor turns on at a higher voltage than the turn-on voltage of the ESD protection device, stacked NMOSFETs (M1 and M3) are sometimes used in the driver. The input node "IN" of the driver is assumed to be grounded during any ESD events (worst case scenario assumed).

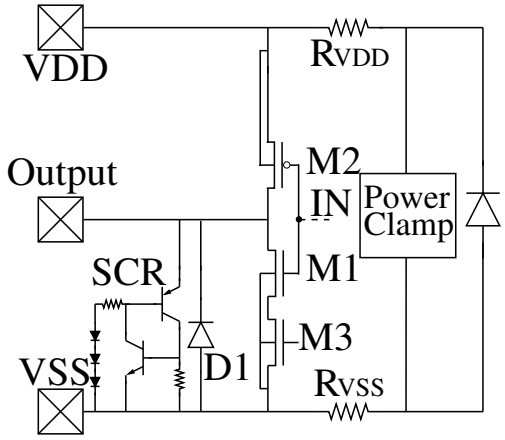
Dual-diode based ESD protection

First, the dual-diode case (Fig. 3.2a) is discussed. Analogous to the receiver with dual-diode protection, PVDD and NVSS are not the limiting case.

During PVSS, the stress V_{ESD} is applied directly on the drain terminal of the PMOSFET (M2) with gate grounded, while the source and body of M2 are coupled high to the value V_{SB} ($< V_{ESD}$) due to the voltage drop on the VDD bus and the supply clamp. Consequently, the GOX of M2 is stressed by a positive pulse applied on



(a) Dual-diode protection



(b) SCR based local protection

Figure 3.2: Simple driver circuit with (a) dual-diode and (b) SCR ESD protection.

the drain as well as the coupled positive voltage (V_{SB}) on the source/body terminals with respect to the grounded gate terminal. For M1 during PVSS, the GOX is stressed by the positive pulse on the drain with substrate and gate grounded while the source terminal is effectively floating (assuming M3 is in the off state to increase the lateral bipolar turn-on voltage).

During NVDD, the GOX of M2 is stressed by a negative pulse on the drain with gate, source and substrate grounded. For the NMOSFET M1 during NVDD, the GOX is stressed by a negative pulse on the drain, with reference to the gate; the

source and substrate terminals are coupled to some negative potential due to the voltage drop on the bus resistance and the supply clamp.

ESD protection with local clamp device

Second, consider the case of local ESD protection (Fig. 3.2b). The stress configurations during NVSS and NVDD are similar to those with dual-diode protection. For PVDD and PVSS, the stress on the GOX of the PMOSFET M2 is also similar to that with the dual-diode protection because of the p+-n diode formed by the drain-body junction.

For the NMOSFET under PVSS, the drain is under positive stress with gate and body grounded. For PVDD, the body of the NMOSFET is pulled high due to the voltage drop on the VSS bus resistance and the supply diode; the NMOSFET GOX is stressed by a positive pulse on the drain, referencing gate with its body terminal coupled to some positive potential.

In summary, GOX of the transistors in the driver is stressed by a pulse on the drain, with reference to the gate, while the other terminals are floating, coupled to some voltage levels or grounded.

3.3 Device Selection

Measurements were conducted on NMOSFETs in 32 nm bulk CMOS technology. All the devices have multiple gate fingers with source and drain fully silicided. The top view of a MOSFET is shown in Fig. 3.3. Devices with the following two different dielectric thicknesses will be explored under ESD-like stress:

1. Thin oxide (SG) devices: $V_{DD} = 1.0$ V and $t_{inv} = 1.46$ nm,
2. Thick oxide (EG) devices: $V_{DD} = 1.5$ V and $t_{inv} = 3.2$ nm.

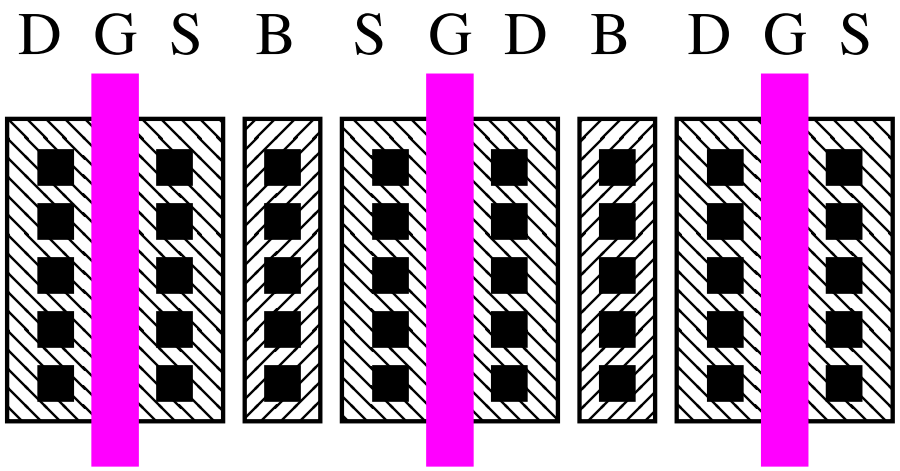


Figure 3.3: Top view of fully silicided MOSFETs. The devices under test have multiple gate fingers.

SG devices are typically used in the core circuits and EG devices are used mainly in input/output circuits. All reported dielectric thicknesses are effective oxide thicknesses t_{inv} measured in inversion:

$$t_{inv} = \frac{\epsilon_0 \epsilon_{SiO2}}{C_{inv}}. \tag{3.1}$$

It is the equivalent oxide thickness if the dielectric permittivity is assumed to equal to that of SiO_2.

3.4 High-k gate dielectric breakdown with stress applied on the gate

In this section, we present and discuss the results of high-k gate dielectric breakdown when the stress is applied on the gate of the MOSFETs. This emulates the stress conditions for the devices in the input receiver.

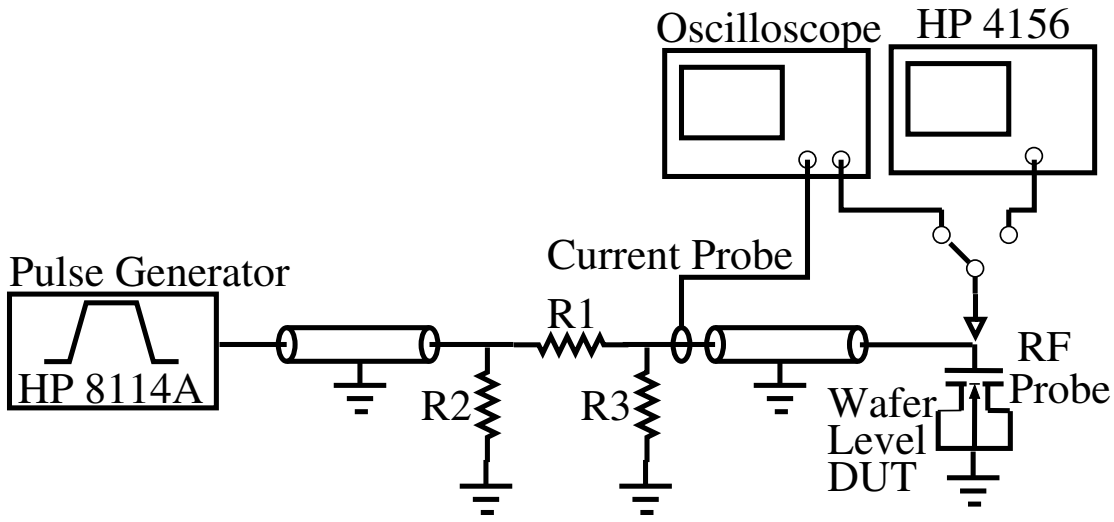


Figure 3.4: Experiment setup for pulsed gate oxide breakdown measurements.

3.4.1 Experimental details

The CVS method is used to determine the breakdown voltage V_{BD} for different ESD events on different time scales (100 ns, 30 ns and 1 ns). The device-under-test (DUT) is stressed by only one pulse. The pulse width is chosen by trial and error so as to ensure that it is long enough so that the gate dielectric under stress fails within one pulse. The same set of experiment is repeated at several stress voltages, V_G, to determine the mathematical relationship between t_{BD} and V_G. Then V_{BD} for a specific t_{BD} can be calculated by either interpolation or extrapolation.

The data acquisition system used for pulsed breakdown measurements is shown in Fig. 3.4. This system is a conventional TLP system with a π resistance network for impedance matching and waveform shaping [78]. The voltage pulse is applied at the gate of the DUT with all other terminals (drain, source and body) grounded. The current of the DUT during the pulsed stress can be measured directly by a current probe (direct method) or it can be calculated by probing the voltages at the two terminals of R_1 (indirect method). The current probe can only measure a current pulse with a pulse width between 30 ns and 1 μs while there is no such limit by

using the indirect method. However, the current waveform obtained by the indirect method turned out to be too noisy. Therefore only the current pulse between 30 ns and 1 μs were captured using the current probe in our experiments. To determine the time-to-breakdown (t_{BD}) of the gate oxide, voltage waveforms were captured by the oscilloscope and saved for analysis. The semiconductor parameter analyzer was used to measure device DC characteristics before and after the pulsed stress.

Since gate oxide breakdown is caused by the generation of defects at random positions inside the gate stack, time-to-breakdown t_{BD} and breakdown voltage V_{BD} are statistical quantities which follow the Weibull distribution (Eq. 2.3). In the following discussions, $t_{63\%}$ and $V_{63\%}$ are used as the characteristic time-to-breakdown and the characteristic breakdown voltage of the DUT, respectively.

3.4.2 Gate oxide breakdown detection

The time-to-breakdown of the gate dielectric can be determined from the captured voltage waveform. After the failure of gate dielectric, the impedance of the DUT changes and this will cause the captured voltage waveform on the DUT to drop due to the load-line effect. Fig. 3.5 illustrates the representative voltage and current waveforms of a high-k gate NMOSFET stressed in accumulation. A significant impedance drop after oxide breakdown is reflected by the voltage drop (and current increase) on the waveform. However, it can also be seen that the voltage drops first and then recovers to its original value. It is important to find out if this feature indicates the failure of the gate dielectric. Several experimental results confirm that the voltage drop and recovery is indeed a failure event.

The first supporting result is shown in Fig. 3.6. Fig. 3.6 shows a representative voltage waveform with reduced stress voltage and increased stress time so that each voltage-drop-recovery event can be clearly identified. The voltage drop and recovery

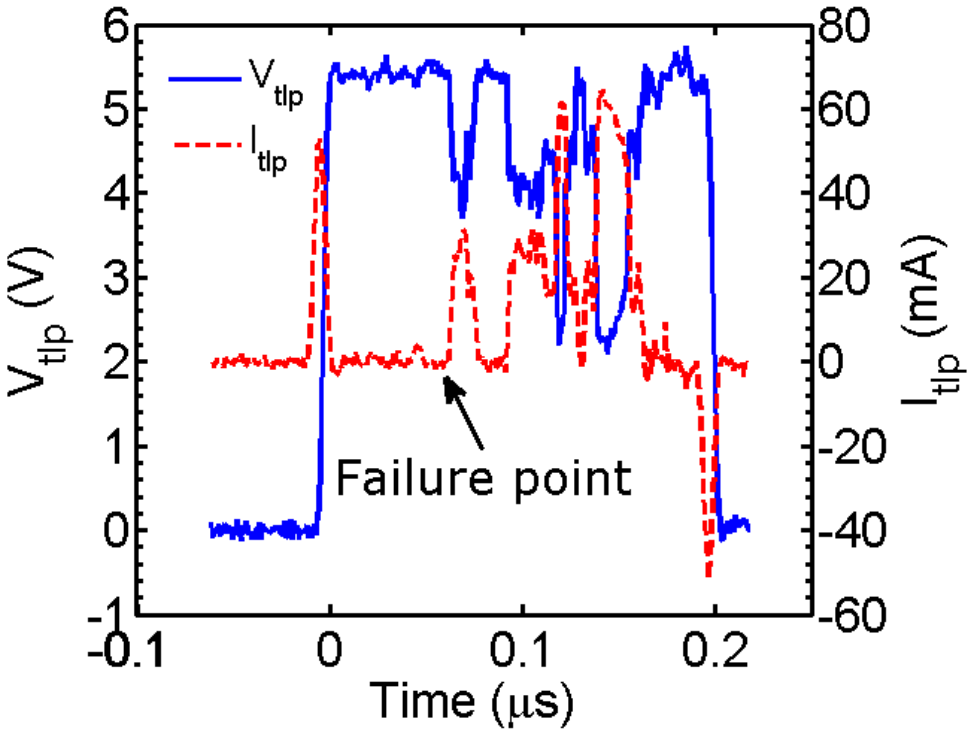


Figure 3.5: Typical gate voltage waveform for an NMOSFET with high-k gate dielectric illustrating the feature of gate dielectric breakdown.

forms multiple spikes on the captured waveform. Interestingly, we noticed that the number of the voltage-drop-recovery events is well correlated with the number of the gate fingers of the DUT. Assume that t_n corresponds to the stress time at which the n^{th} voltage drop occurs. The cumulative distributions of t_1, t_2 and t_3 etc., as shown in Fig. 3.7, can be obtained by stressing multiple devices. The x axis is the normalized time $\mu = (t_n/t_{63\%})^{\beta}$, where $t_{63\%}$ and *beta* are obtained by fitting t_1 to the Weibull statics (Eq. (2.3)). It turns out that the statistical distribution of t_1, t_2 and t_3 etc. can be modeled by the successive breakdown statistics:

$$\ln\left[-\ln\left(1-F\right)\right] = \ln\left\{\mu - \ln\left(\sum_{i=0}^{n-1}\frac{\mu^i}{i!}\right)\right\} \tag{3.2}$$

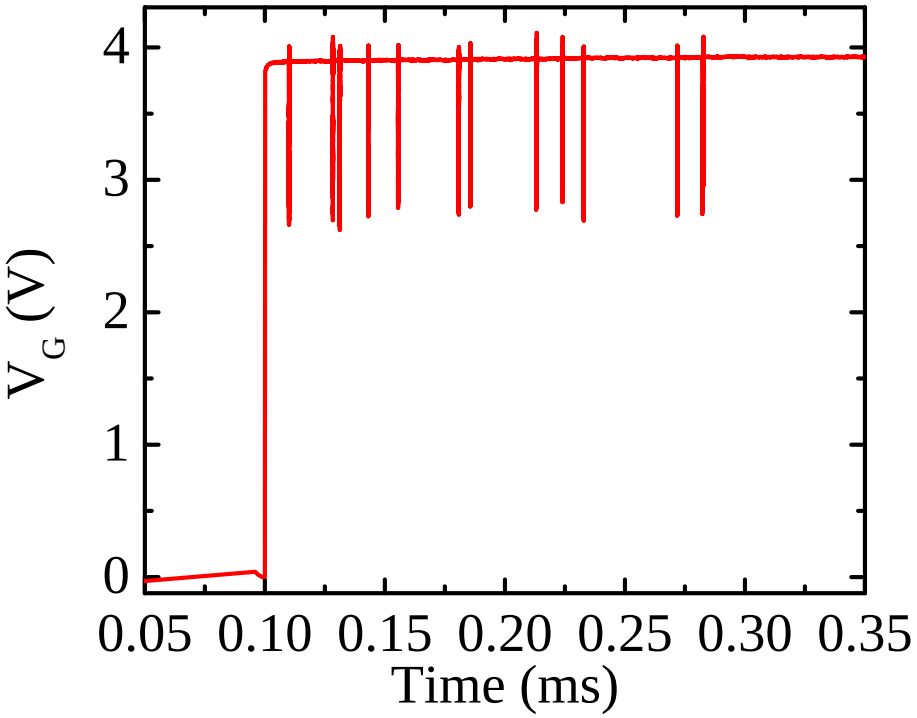


Figure 3.6: Voltage drop and then recovery forms multiple spikes on the captured voltage waveform. The number of spikes is well correlated with the number of gate fingers.

as described in [79], where F is the cumulative rate of the n^{th} voltage drop. The solid line in this figure is calculated based on Eq. (3.2). The excellent agreement between the experimental data and the theoretical model indicates that these voltage spikes are independent of each other. The above results suggest that each of these voltage spikes corresponds to the failure of a single gate finger. The current flowing through the device after oxide failure can be very large (larger than 20 mA) and this high current density may cause a loss of contact to the damaged gate finger, which increases the gate oxide impedance to the value similar to that before breakdown. This leads to the "recovery" of voltage in the waveform.

The second supporting experimental results is from the area scaling. As discussed

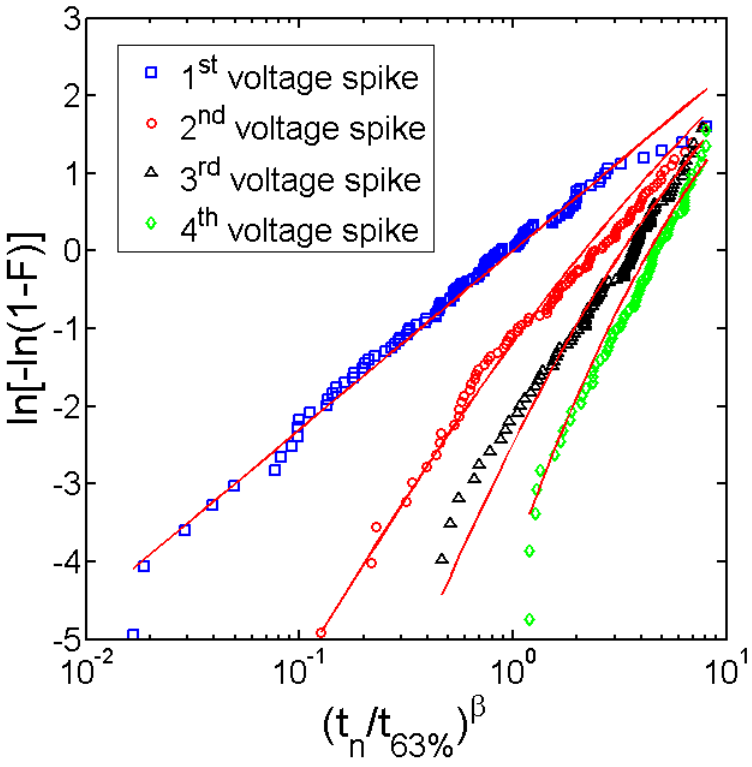


Figure 3.7: The cumulative distributions of t_1, t_2 and t_3, where t_n represents the time when the nth voltage drop occurs.

in Chapter 2, gate oxide breakdown is due to the percolation path created by the random defect generation. Larger area makes it easier to form the percolation path and hence the time-to-breakdown will be shorter. Assuming the voltage spike is the failure of one gate finger, the time it takes to observe these spikes should be equal to the failure time of the gate dielectric with the area equal to the area of one gate finger, which is 0.05μm^2. Also the failure time of gate dielectric with larger areas can be taken as the time of the first voltage spike. These data should follow the area scaling law (Eq. (2.6)). Fig. 3.8 shows that the area scaling law is followed very well. This again confirms that the voltage spikes observed in Fig. 3.6 indicates the failure of gate fingers.

Gate dielectric breakdown can be caused by either stress-induced random defect

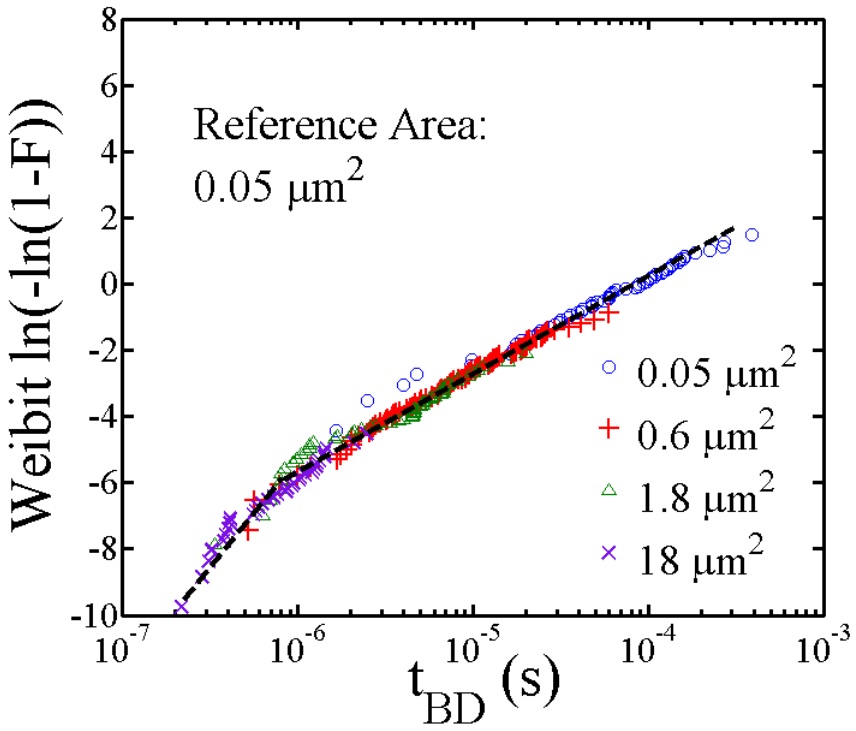


Figure 3.8: Gate oxide failure follows the area scaling law, indicating that the breakdown follows the intrinsic breakdown process.

generation inside the gate stack or extrinsic defects introduced during the various process steps. The area scaling property shown in Fig. 3.8 reveals that the dielectric breakdown here is related to the intrinsic breakdown process and it is not governed by the extrinsic defects inside the gate dielectric. It is also interesting to notice that the slope of the Weibull distribution increases from the high failure percentile (corresponding to smaller gate oxide area) to the low failure percentile (corresponding to larger gate oxide area). Similar observations were also made on SiON and high-k gate dielectrics under low stress voltages [75,80,81]. For the high-k dielectric, this may be attributed to a much higher defect generation rate in the high-k layer compared to that in the interfacial layer [75, 81], while the progressiveness of the breakdown path wearout is believed to cause the slope change in SiON dielectrics [80].

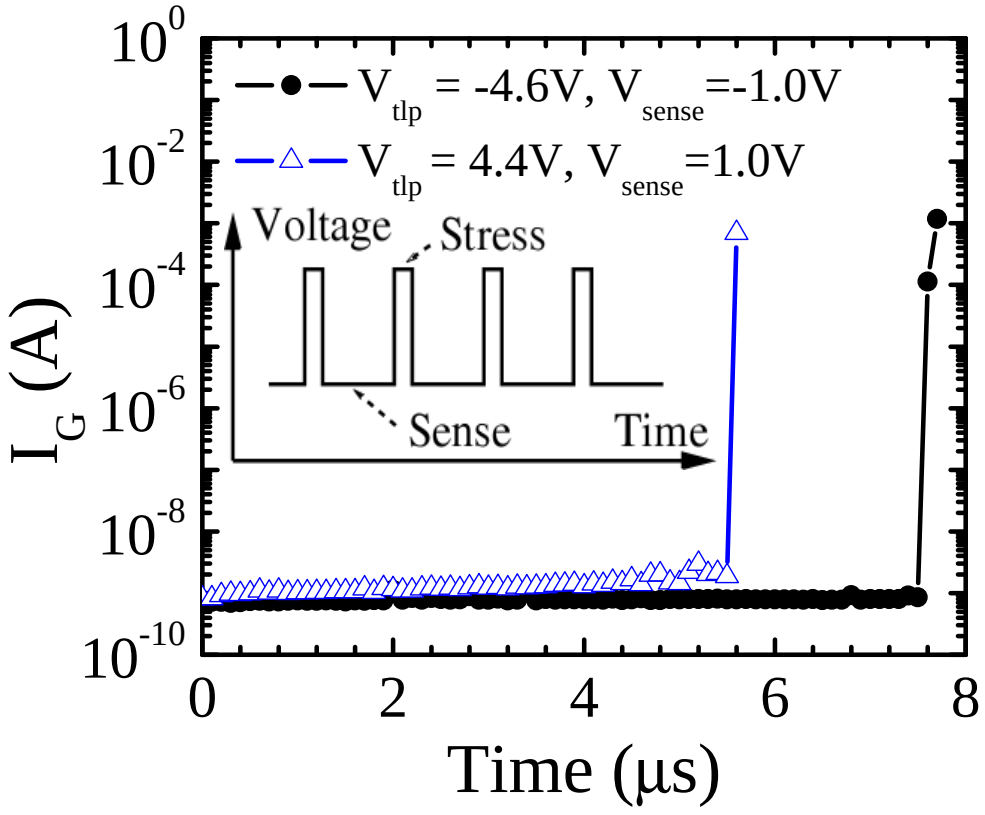


Figure 3.9: RCVS shows that gate leakage current changes slightly before the hard failure. The breakdown features an abrupt increase in I_G.

3.4.3 Gate oxide failure: hard or soft?

Due to the limited resolution of the current probe, the minimum current that can be measured during a TLP pulse is 1 mA. Consequently, detection of soft breakdown, which features a small increase in gate leakage current (I_G) magnitude or noise, is not possible. Additionally, recent studies [55] showed that the breakdown transient for stress voltages larger than 4 V or for devices with metal gate can be extremely fast. This also indicates that soft breakdown is generally not relevant to this study, which used stress voltages larger than 4 V in most cases and DUTs with metal gate electrodes. The gate oxide breakdown reported in this study is catastrophic.

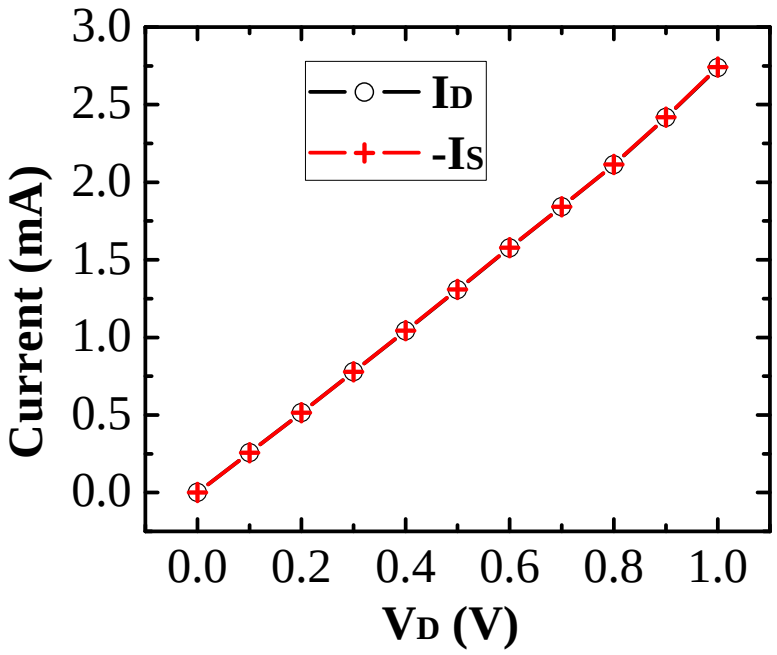


Figure 3.10: The drain to source current after gate oxide breakdown under either positive or negative stresses showed a resistor-like characteristics for devices with different channel lengths (SG: 50 nm and EG: 150 nm)

The conclusion above is also supported by the following experiment results. First, the failure features an abrupt increase of I_G in both stress polarities as shown by Fig. 3.9. The gate leakage current does not change much after the voltage "recovery". In order to measure the gate leakage current immediately after oxide breakdown (before the "recovery"), the gate oxide was stressed by a series of pulses with constant amplitude, i.e. Repetitive Constant Voltage Stress (RCVS), as shown by the inset of Fig. 3.9. Since the width of each pulse is shorter (100 ns), there is a higher chance that the stress is stopped before the "recovery". As shown in Fig. 3.9, the gate leakage current changes slightly first and then increases abruptly which corresponds to the leakage current measured immediately after the pulse with a voltage drop on the captured waveform. In PMOSFETs, the breakdown results show similar characteristics.

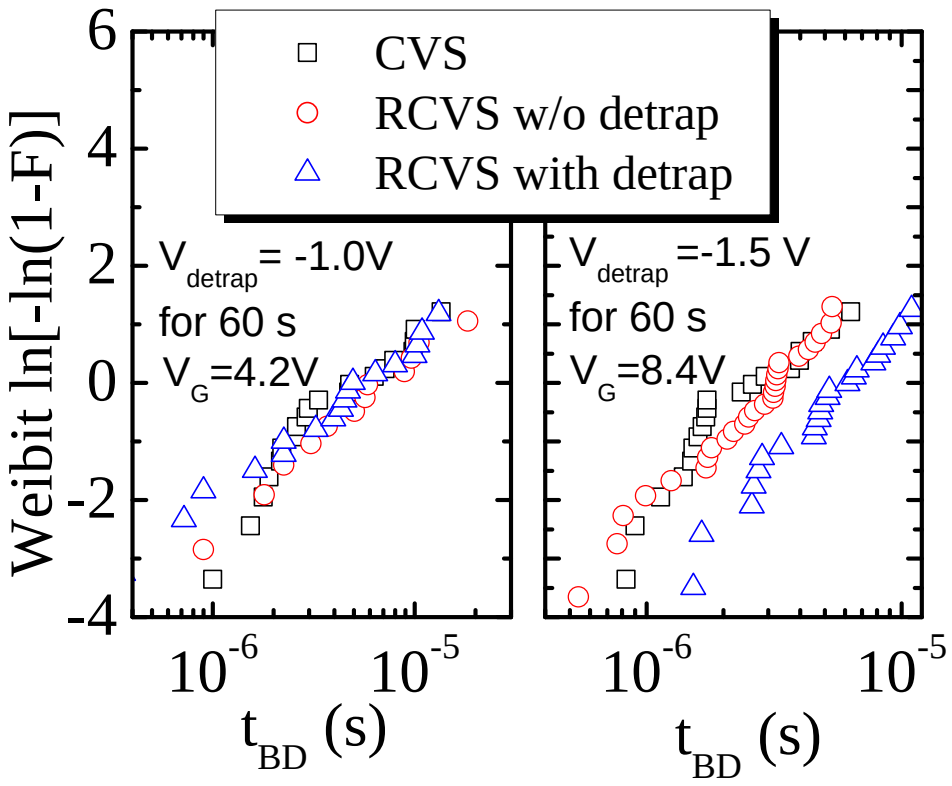


Figure 3.11: Effect of stress interruption and charge detrapping on the t_{BD} of thin (left) and thick (right) oxide devices.

Second, the drain to source current after gate oxide breakdown under either positive or negative stresses shows a resistor-like characteristics for devices with different channel lengths, as shown in Fig. 3.10. This may be caused by electromigration due to the high electric field and the large temperature gradient during the gate dielectric breakdown transient.

3.4.4 Effect of stress interruption on high-k gate dielectric breakdown

High-k dielectric contains a large amount of electron traps, presumably oxygen vacancies [13]. Electrons get trapped in the gate oxide during the positive stress and

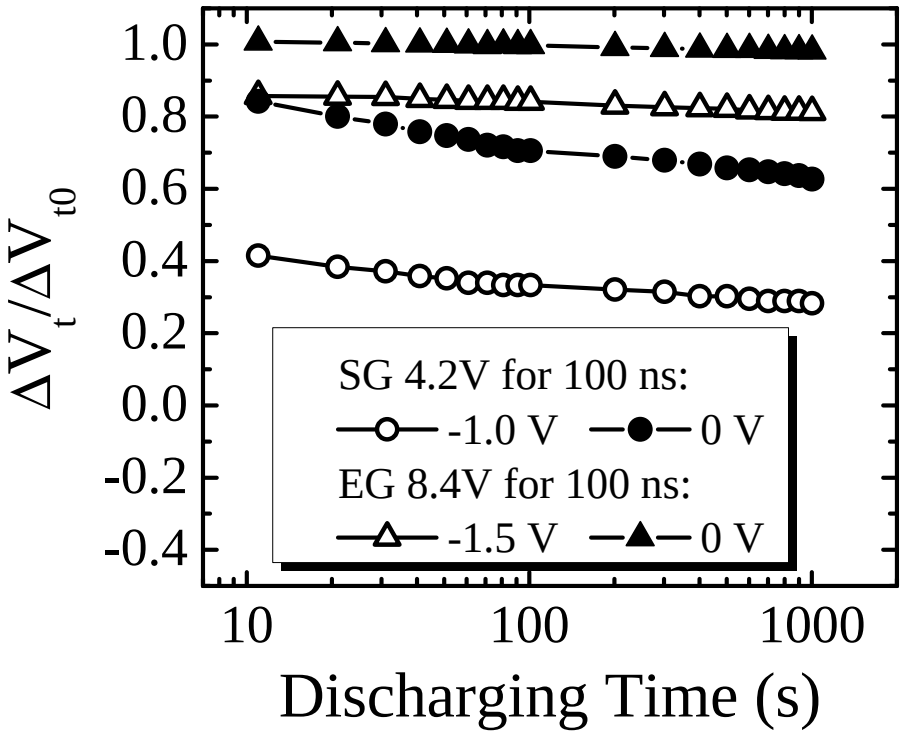


Figure 3.12: Recovery of threshold voltage at different gate voltages indicates that the trapped electrons inside the gate stack were detrapped during the period when the stress was interrupted.

detrapped if the stress is interrupted. This may affect the time-to-breakdown distribution as they change the electric field distribution inside the gate stack. To study the effects of stress interruption on t_{BD}, the time-to-breakdown distributions measured from CVS and RCVS are compared as shown in Fig. 3.11. For one group of devices under RCVS, the stress was interrupted periodically only for DC leakage measurements; for the other group of devices under RCVS, a small negative bias was applied on the gate for 60 s between the stress pulses. Interruption of the stress resulted in electron detrapping from the gate stack; a negative bias on the gate helped accelerate the detrapping process. To avoid any damage caused by the negative bias, -1 V was used for SG devices and -1.5 V for EG devices. Fig. 3.11 shows that interrupting the

stress ($\sim$ 5 s) for DC leakage measurement did not change the gate oxide time-to-breakdown distribution of both SG and EG devices compared to CVS measurement. However, the negative bias between stress pulses increased the t_{BD} of EG devices by a factor of two. The change of the device threshold voltage reflects the amount of trapped electrons. Fig. 3.12 shows the influence of interrupting the stress ($V_G = 0$ V) and the influence of the explicit negative discharging bias on the device threshold voltage. The explicit discharging process helped remove a considerable portion of the trapped electrons. A much bigger portion of the trapped charges was removed by the negative bias in SG devices compared to EG devices, but surprisingly, the discharging process did not change the SG t_{BD} distribution, whereas it increased the EG t_{BD}. This indicates that the time-to-breakdown of the thick high-k gate is more sensitive to the trapped electrons in the gate stack.

3.4.5 Time-to-breakdown of high-k gate dielectric

Fig. 3.13 shows $t_{63\%}$ obtained using the CVS method as a function of gate voltage for NMOSFETs and PMOSFETs with thin gate oxide (SG) in both inversion and accumulation. For $t_{63\%} > 1$ s, the data was obtained using the semiconductor parameter analyzer in the sampling mode. Time-to-breakdown in this case is defined as the time at which the gate current at the stress voltage increases abruptly by at least 10 μA. In order to predict V_{BD} for different ESD events (with different durations), V_{BD} should be expressed as a function of $t_{63\%}$. Fig. 3.13 shows that the data $t_{63\%}$ can be fitted well by the power law [82]:

$$t_{63\%} = c \cdot V_G^{-n} \tag{3.3}$$

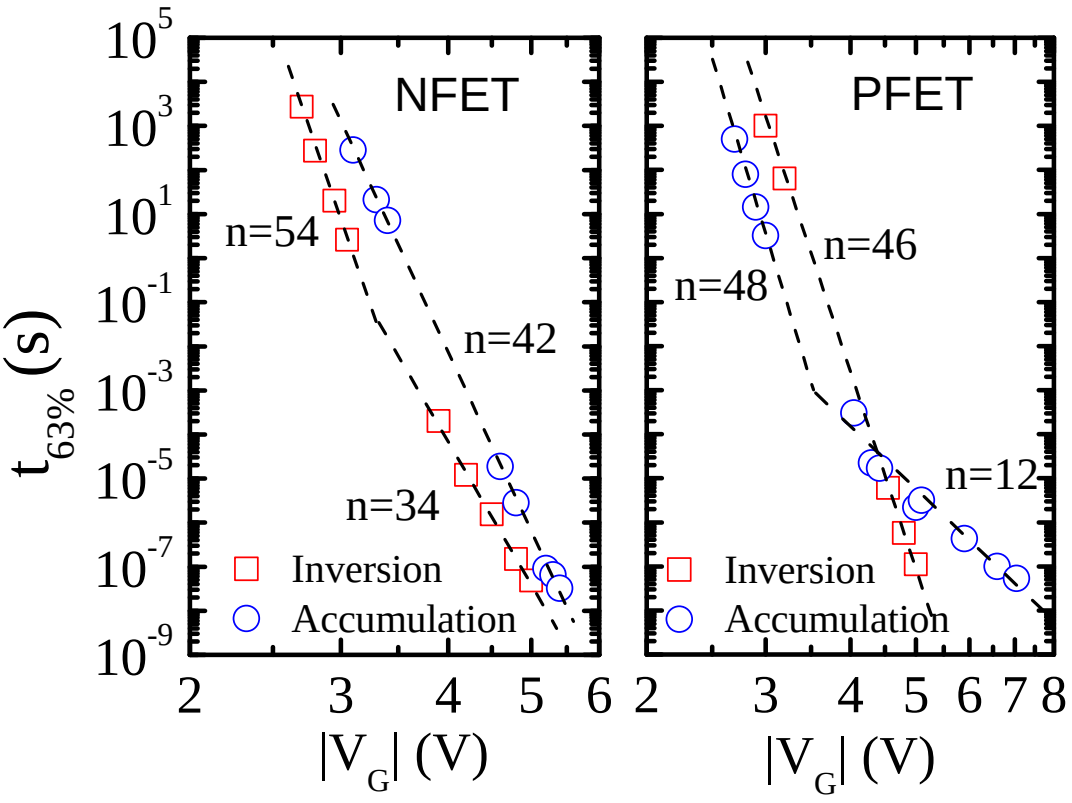


Figure 3.13: Time-to-breakdown of thin-oxide MOSFETs with high-k gate dielectrics, obtained using the CVS method.

over a limited time range. For NMOSFETs and PMOSFETs stressed negatively, the power law exponent n stays constant in the entire time range evaluated in this study. However, the power law exponent for PMOSFETs under positive stress (accumulation) decreases considerably as the stress voltage increases.

This significant change of n in PMOSFETs is due to the effect of well resistance. For devices under positive stress, the majority of the gate current is coming from the tunneling of conduction band electrons. In PMOSFETs, the electron is provided by the N-well through the body contact. Fig. 3.14 shows that the majority of I_G flows through the N-well to ground. As the tunnling leakage current increases exponentially with the gate voltage, a significant portion of the stress voltage drops across the well

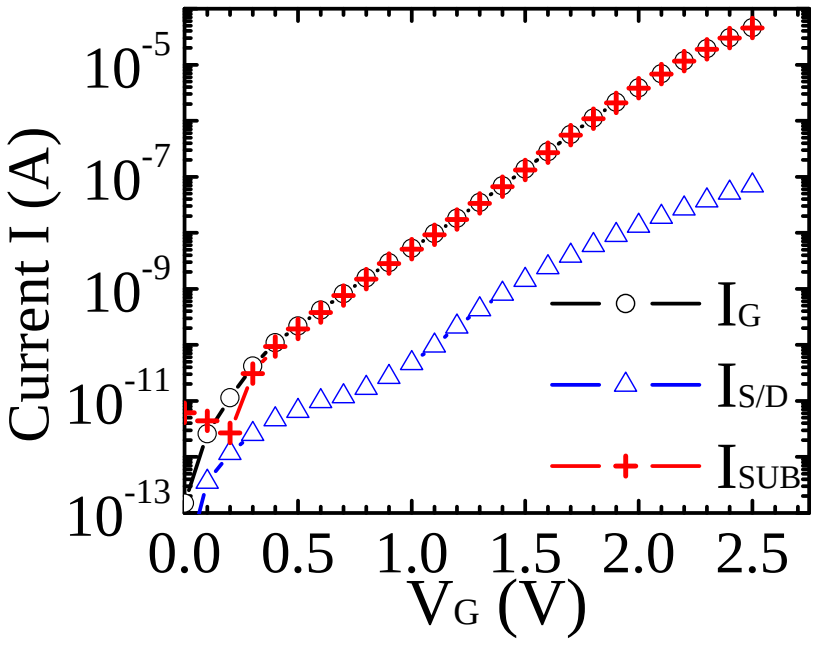


Figure 3.14: Terminal currents measured on a PMOSFET in accumulation (positive stress).

resistance and the effective stress on the gate oxide is reduced. This bends the t_{BD} vs. V_G curve and decreases the power law exponent. To verify the above argument, devices with two different distances between the gate edge and the body contact (i.e. different well resistance) were measured. As illustrated by Fig. 3.15, for a given V_G, the $t_{63\%}$ of the transistors with a larger well resistance is about 10 times longer. For NMOSFET stressed in accumulation, the body current is made of holes. As the tunneling barrier height for holes is much larger compared to electrons, the body current in this case is not large enough to cause the IR effect. Therefore, the voltage drop over the well resistance is not sufficient to affect the $t_{63\%}$ of NMOSFETs in accumulation (under negative stress). Although a small reduction of the power law exponent was also observed for NMOSFETs under positive stress (inversion), this is not a general trend as a constant power law exponent was observed for $t_{63\%}$ in the

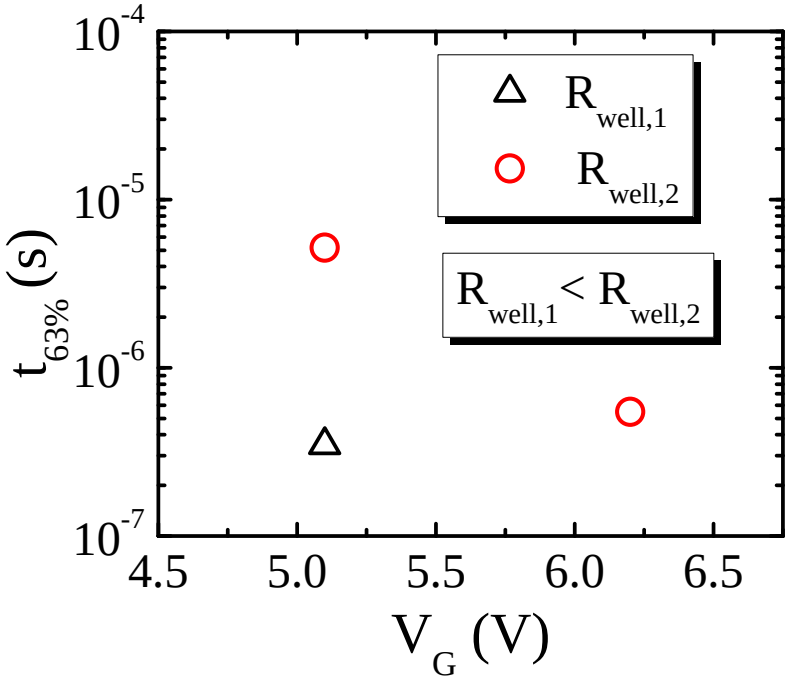


Figure 3.15: The effect of well resistance on $t_{63\%}$ of PMOSFETs in accumulation.

range from 100 ns to 100 s on some of the wafers.

In SiO_2 gate, the charge-to-breakdown ($Q_{BD} = J_G \cdot t_{BD}$) measured in gate injection mode (negative stress) is smaller than that measured in substrate injection mode (positive stress) [82]. The DC measurement on high-k gate shows that the charge to breakdown under positive and negative stress is similar. However, the gate leakage current for positive bias is much larger than that for negative bias due to the asymmetry in the high-k gate stack. Considering this, the $t_{63\%}$ is smaller when stressed by positive voltages than by negative stresses for a given voltage magnitude, as shown by the $t_{63\%}$ of the NMOSFETs and that of the PMOSFETs under small stress voltages in Fig. 3.13. Because of the well resistance, the trend is reversed for PMOSFETs at large stress voltages. As a result, NMOSFET under positive stress is found to have the smallest breakdown voltage.

As discussed in Sec. 3.2, the GOX may be stressed in inversion when the source

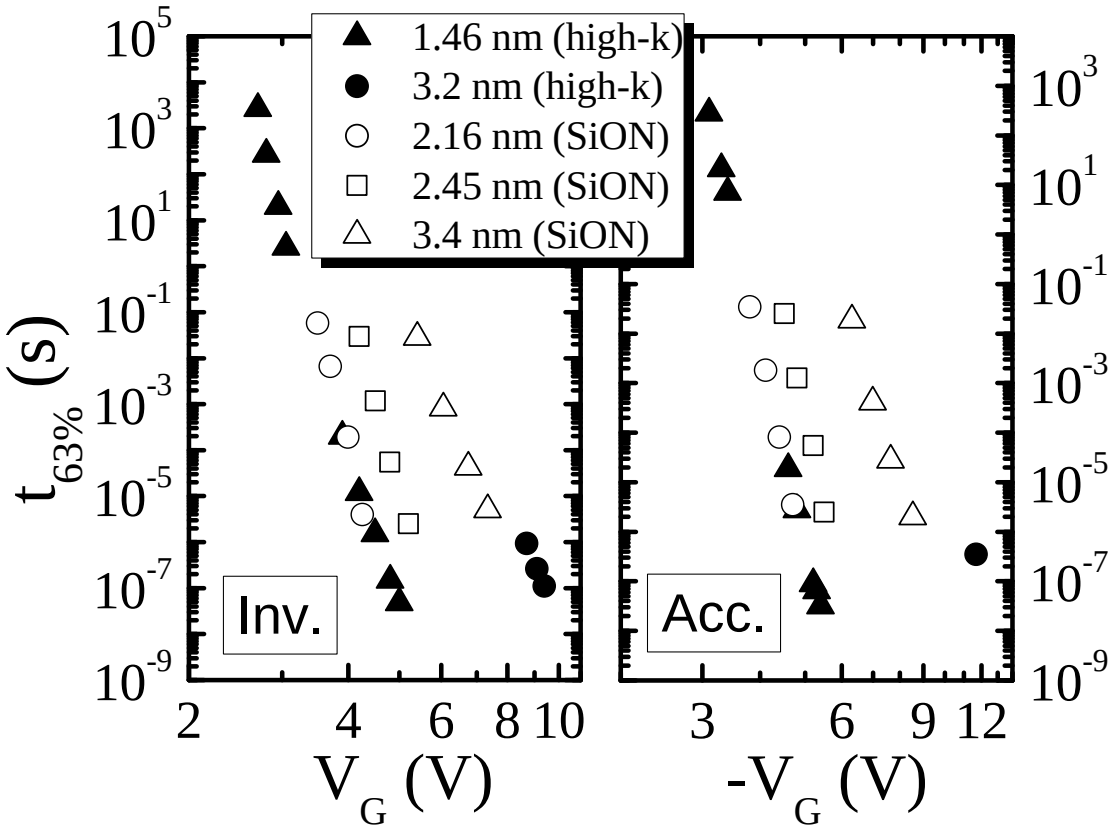


Figure 3.16: Comparison of $t_{63\%}$ between high-k and SiON gate dielectrics.

and body are grounded with drain floating. Experimental results for testing under this condition show that floating the drain does not affect the GOX breakdown voltage much. This is because the channel length is so short (50 nm) and floating the drain has thus little impact on the resistance of the conduction path or the area of the GOX under stress.

3.4.6 Comparison of high-k gate dielectric with SiON gate dielectric

Fig. 3.16 compares the $t_{63\%}$ of NMOSFETs with high-k gate and SiON gate under positive and negative stress. Devices with SiON gate are from a 45 nm bulk CMOS

technology. Although t_{inv} of the thin-oxide high-k NMOSFETs is much smaller than that of the thin-oxide SiON device (1.46 nm compared to 2.16 nm), $t_{63\%}$ of these two devices are comparable. The result suggests that for a given effective oxide thickness, high-k gate have a larger breakdown voltage and hence increased ESD robustness compared to SiON gate, despite its smaller ultimate breakdown strength E_{BD} (electric field at breakdown) as suggested by the thermochemical model [3]. E_{BD} of a dielectric material is found to decrease as the dielectric constant ϵ increases; it is proportional to $\epsilon^{-0.65}$ [3]. However, for the same effective oxide thickness, the physical thickness of the high-k gate is $\epsilon_{HK}/\epsilon_{SiON}$ times larger than that of the SiON gate. Therefore, for the same effective oxide thickness, the breakdown voltage of the high-k gate ($E_{BD} \times t_{ox,physical}$) is still larger.

3.5 High-k gate dielectric breakdown with stress applied on the drain

Sec. 3.4 discusses the situation when the stress is applied on the gate of the MOSFETs, which emulates the MOSFETs under stress in the input receiver circuit. Besides this stress configuration, Sec. 3.2 shows that for the MOSFETs in the output driver, the stress is applied on the drain of the MOSFETs. During the ESD event, the GOX stress may be stressed on the drain terminal with respect to the gate terminal, while the other terminals floating, coupled to some voltage levels or grounded. In this section, this stress configuration is characterized using the TLP method [1].

First, we discuss the case when the source is floating. Fig. 3.17 shows the GOX breakdown voltage when the drain of the NMOSFET is under positive stress with the gate grounded and the source/body floating. The results are compared with

[1] Comparison between TLP and CVS will be discussed in the next section.

V_{BD} when negative stress is applied on the gate with source/drain/body grounded. Measurements were performed on both dual-well (DW) and triple-well (TW) test structures. The p-type body of the TW device is isolated from the p-substrate by an n-well, while the p-type body of the DW device is resistively connected to the p-type substrate. Although there is no test probe directly touching the body contact of the DW transistor during the stress test, the body is virtually grounded because of the large wafer substrate and the chuck of the probe station. As a result, the potential at the oxide/silicon interface is controlled by the grounded gate and essentially very small. Only the GOX in the drain/gate overlap region is stressed. This is confirmed by an experiment in which the body of the DW transistor was grounded explicitly; no difference in the V_{BD} was observed compared to the case with the "floating" transistor body. Since GOX with a smaller area has a larger V_{BD} V_{BD} is significantly increased for the DW transistors with this stress configuration, in which only the small drain-gate overlap region is stressed. For the TW transistor, the transistor body is more isolated from the large p-type substrate compared to the DW devices. During the test, the capacitive coupling between the drain and the body raises the body potential more easily than in DW devices. This increases the area of the GOX under stress and hence reduces the V_{BD} as shown in Fig. 3.17.

The result in Fig. 3.17 indicates that when the drain of the NMOSFET is stressed positively with the gate grounded, the breakdown voltage of the GOX should be between V_{BD1} and V_{BD2}. V_{BD1} ($< V_{BD2}$) is obtained by stressing the gate negatively with all other terminals grounded, while V_{BD2} is obtained by stressing the drain positively with the gate and body terminals grounded and the source terminal floating. In other words, V_{BD1} is the breakdown voltage when the entire GOX is under stress, while V_{BD2} is the breakdown voltage when the drain/gate overlap region is under stress. The exact breakdown voltage when the stress is applied on the drain terminal

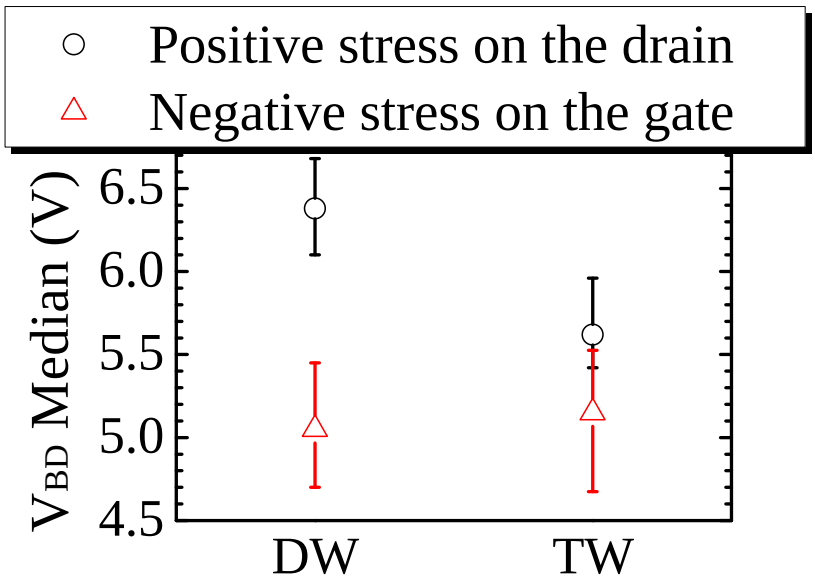


Figure 3.17: Gate dielectric V_{BD} for positive stress on the drain with gate grounded and substrate/source floating, in comparison with the VBD for negative stress on the gate with other terminals grounded.

with source floating is determined by the voltage coupling between the drain and the body. Higher coupling increases the gate dielectric area under stress and hence decrease the breakdown voltage.

If the device is stressed on the drain with all other terminals grounded, the parasitic bipolar transistor turns on before the GOX breakdown is reached in these fully silicided devices, resulting in immediate failure of the transistors due to current crowding.

3.6 Comparison between CVS and TLP method

The CVS results were compared with the results obtained by the conventional TLP method which is widely used in the characterization of devices under conditions similar to those of an ESD event. Unlike the CVS method, the TLP method applies

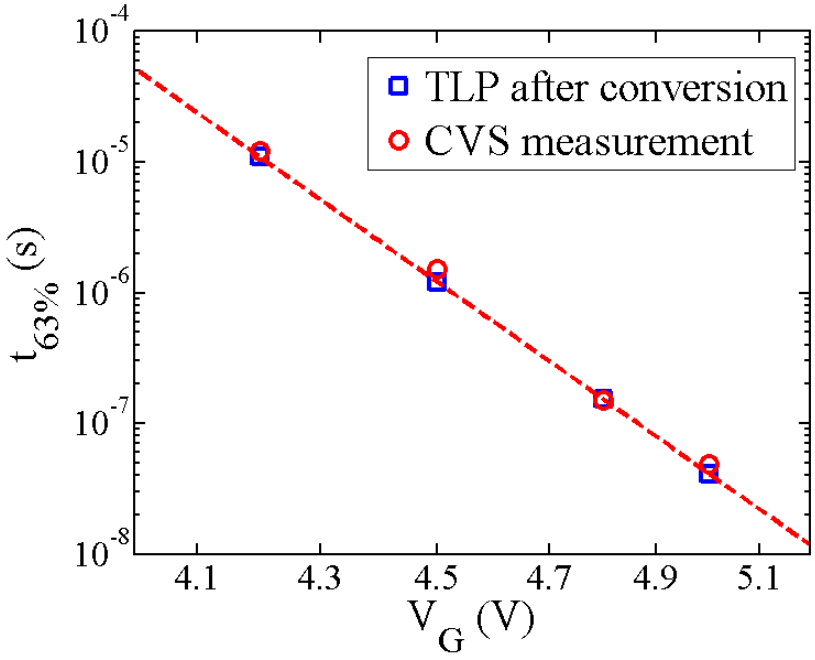


Figure 3.18: Data collected using the TLP method can be converted to $t_{63\%}$, which is comparable to $t_{63\%}$ measured using the CVS method.

a series of pulses with successively increasing amplitudes until a failure is detected. The breakdown voltage V_{BD} is measured directly and the corresponding time-to-breakdown can be taken as the applied pulse width t_{pw}. Since the multiple pulses applied during a TLP test introduce error caused by the cumulative damage, V_{BD} obtained from TLP is expected to be smaller than that from CVS. However, it is desirable to use the TLP method which is more convenient and efficient than CVS.

During TLP, the magnitude of the voltage pulse with a fixed width (t_{pw}) is ramped up with an increment, ΔV, in each step until the gate oxide breakdown is detected. The breakdown voltage corresponding to the breakdown time $t_{BD} = t_{pw}$ can be acquired directly from the measurement. The cumulative damage from the multiple stresses has to be taken into account to get the exact t_{BD} or V_{BD}. Following the procedure proposed in [53], V_{BD} data acquired during a TLP test can be converted

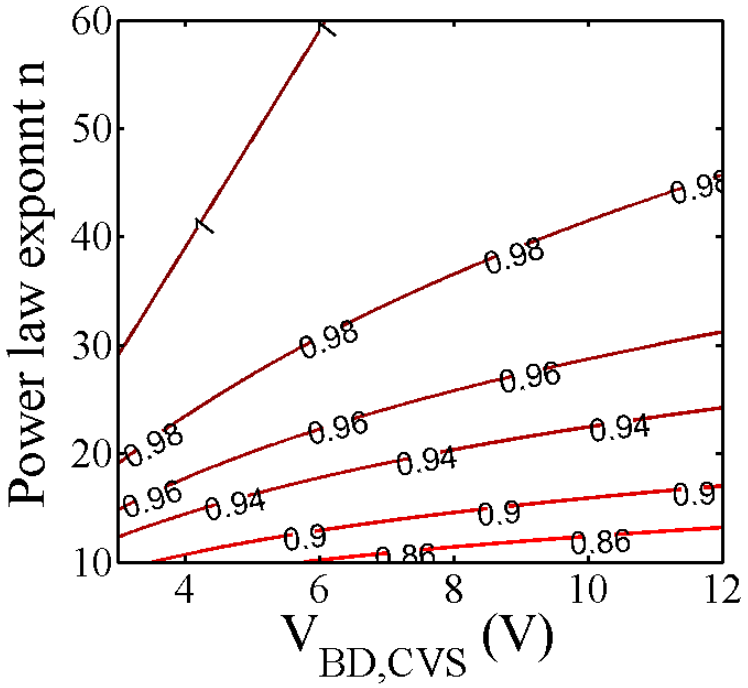


Figure 3.19: Contour plot of $V_{BD,TLP}/V_{BD,CVS}$ as a function of power law exponent (n) and the breakdown voltage from CVS ($V_{BD,CVS}$) with TLP voltage increment step $\Delta V = 0.1$ V.

to $t_{63\%}$ using Eq. (A.1) in Appendix. The converted $t_{63\%}$ data agree with CVS measurements very well, as shown by Fig. 3.18. However, this conversion requires the power law exponent, which is extracted using the CVS method. Time-to-breakdown is very sensitive to the stress voltage, since the power law exponent is typically in the range from 30 to 50. Therefore, only the last few voltage steps in TLP method make significant contributions to the final gate oxide failure. This indicates that the V_{BD} directly from the TLP method should not deviate from the exact breakdown voltage (from CVS method) too much. Fig. 3.19 compares the breakdown voltage from the TLP method, $V_{BD,TLP}$, and from the CVS method, $V_{BD,CVS}$, by a contour plot of $V_{BD,TLP}/V_{BD,CVS}$ within the typical range of two parameters, the power law exponent n and $V_{BD,CVS}$. $V_{BD,TLP}/V_{BD,CVS}$ is calculated following the procedure described in

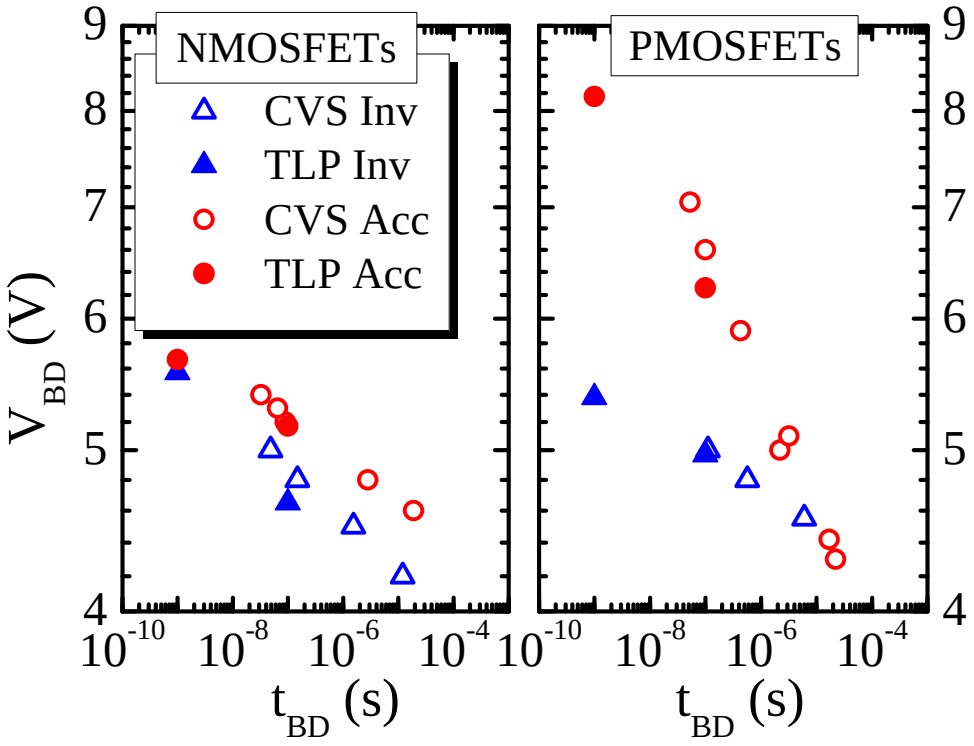


Figure 3.20: Comparison between $V_{BD,CVS}$ and $V_{BD,TLP}$ experiment data. $V_{BD,CVS}$ is directly from Fig. 3.13. t_{BD} from the TLP method is the pulse width t_{pw}.

Appendix. The voltage increment ΔV in the TLP method is assumed to be 0.1 V. It is shown that $V_{BD,TLP}$ correlates well with $V_{BD,CVS}$ in the typical parameter range, and in most cases the difference is less than 10%. Fig. 3.20 compares the experimental $V_{BD,TLP}$ and $V_{BD,CVS}$ measured on thin-oxide MOSFETs. They correspond to the failure percentile $F = 63.2\%$. $V_{BD,CVS}$ is taken directly from the data in Fig. 3.13. As expected, V_{BD} data from both methods are fairly close, with the difference less than 10%. This suggests that V_{BD} from the TLP method can be used as a conservative estimation of the breakdown voltage.

TLP method was also used to measure V_{BD} down to $t_{BD} = 1$ ns on a commercially available very fast TLP system. The results are shown in Fig. 3.20. The 1-ns data stays close to the line extrapolated from the data measured using the CVS method.

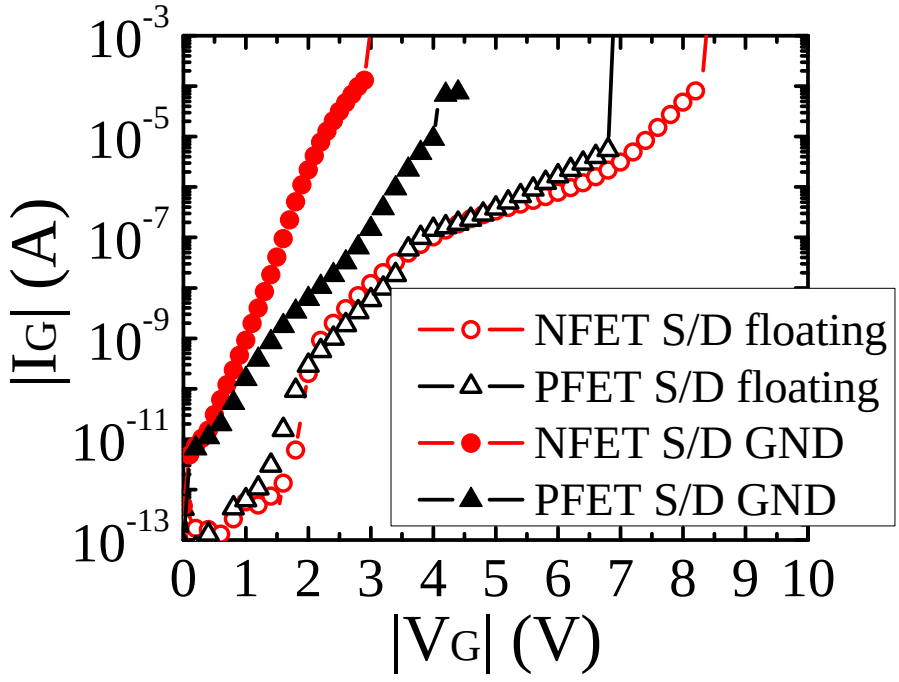


Figure 3.21: DC gate leakage current of NFET and PFET in inversion with source/drain grounded or with source/drain floating.

This suggests that the breakdown mechanism remains the same down to 1 ns range, which is relevant to the CDM-ESD event.

3.7 Method to improve the circuit immunity to GOX breakdown induced failure

As described in Sec. 3.4, the well resistance increases the time-to-breakdown of the GOX for PMOSFETs stressed positively. This indicates that the GOX V_{BD} can be increased by increasing the resistance of the current flowing path, since the voltage drop across the resistor reduces the stress applied on the gate dielectric.

In this section, we show first that the source/drain resistance can help increase V_{BD} of the gate dielectric when the MOSFETs are stressed in inversion. When stressed in

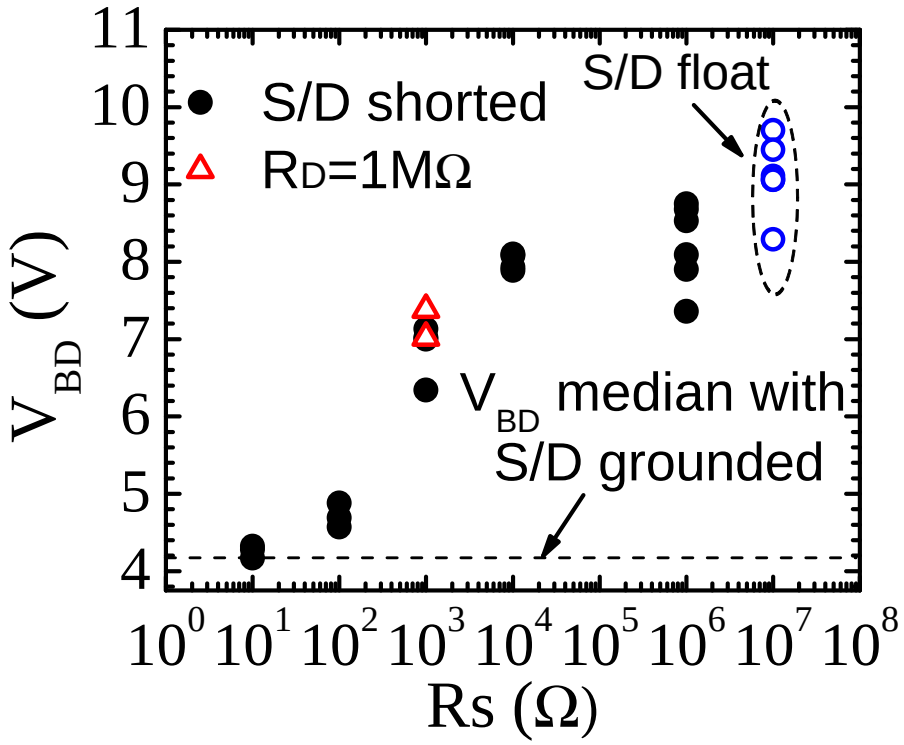


Figure 3.22: Breakdown voltage of NFET in inversion is increased when source/drain resistance is increased.

inversion, the inversion layer charge carriers (electrons or holes) are provided mainly by the source and the drain. If the resistance connected to source/drain is large enough, the charge carriers have to be provided by thermal generation in the transistor body. Therefore, increasing the resistance connected between source/drain contacts and the ground will impede the formation of inversion layer if the stress duration is shorter than the carrier generation lifetime. The transistor under stress will be forced into deep depletion. This reduces not only the voltage drop across the GOX but also the charge flow rate. Consequently, V_{BD} of the GOX is increased. The above argument is supported by the DC gate leakage measurements, as shown in Fig. 3.21 and the VBD data obtained using the TLP method as shown in Fig. 3.22. Fig. 3.21 shows that floating the source/drain contacts can reduce the gate leakage

current and increase V_{BD} of the gate dielectric. Fig. 3.22 shows that when the source and the drain are shorted and connected to ground through a resistor, even a small resistance ($\sim 1000\ \Omega$) is enough to increase the V_{BD} by 50% under TLP stress. It also demonstrates that the V_{BD} is increased when the source and drain are connected to ground through different resistors. For transistors with larger gate dielectric areas, the impact of source/drain resistance on V_{BD} is expected to be more significant because of the larger gate current flowing through source/drain to the resistor and then to ground.

The above discussion suggests that the ESD design margin may be increased by modifying the topology of the I/O circuit. By placing an extra resistive device (resistor or active device with high resistance) in series with the current path through GOX, the pad voltage tolerance or design window can be increased for transistors in both receiver and driver circuits.

A representative design is shown in Fig. 3.23. Two extra transistors, M3 and M4, are added. A similar idea has also been proposed in [83, 84] for the 90 nm and 130 nm technology nodes. Under normal conditions, M3 and M4 are biased in inversion (on-state) to avoid disturbing the normal circuit operation. During ESD events, they are biased in either accumulation or depletion (off-state) and act like resistors which help increase V_{BD} of the GOX as demonstrated in Fig. 8. The gate bias signal for M3 and M4 can be taken from the RC triggered power clamp as shown in Fig. 9. Using these design techniques, the maximum allowable pad voltage is expected to increase considerably. The above strategy requires that the resistance of the supply clamp is small enough to avoid elevating (lowering) the source and the body potential of M4 (M3) too much during positive (negative) discharge to VSS (VDD). Otherwise, the raised (lowered) potential at the source/body terminals of M4 (M3) may over stress the GOX of M4 (M3) and hence the maximum pad voltage may be limited by the

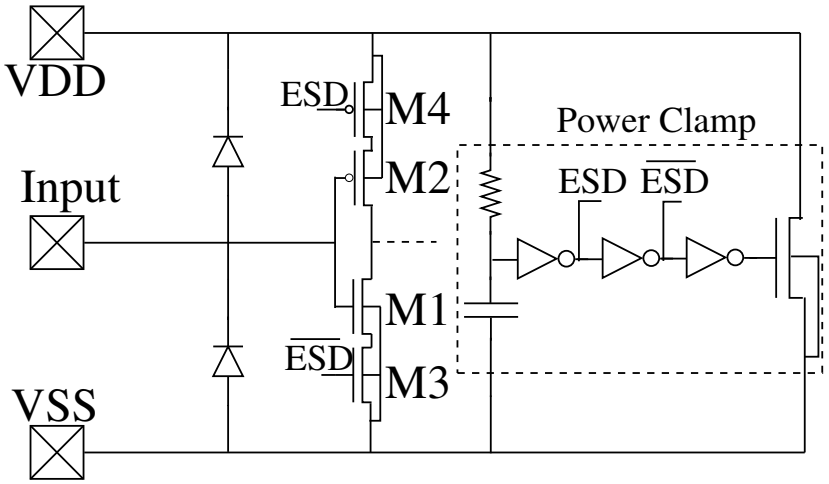


Figure 3.23: Additional transistors M3 and M4 are used so as to increase the breakdown voltage of NMOSFET and PMOSFET in inversion.

GOX failure of M4 (M3).

3.8 Summary

The stress configurations of gate dielectric in the context of real input/output circuits are discussed first. The stress experiments are designed taking into account these different stress condition. We first studied the stress condition for transistors in input receiver in Sec.3.4. It is found that high-k gate oxide breakdown of MOSFETs with high-k gate under ESD-like pulsed stress is catastrophic. The high voltage stress changes the gate leakage current only slightly until the final hard breakdown. After gate oxide failure, the large gate current may result in a loss of gate contact, and the drain-to-source current showed a resistor-like behavior. Interrupting the stress briefly does not change the t_{BD} statistical distribution. The time-to-breakdown of the transistors with thick high-k dielectric is more sensitive to the negative discharge biasing compared to those with thin high-k dielectric. It is found that NMOSFETs under positive stress have the smallest V_{BD}. For PMOSFETs under positive stress

the voltage drop on the well resistance increases the breakdown voltage significantly. The comparison between MOSFETs with high-k gate and SiON gate showed that for a given effective gate oxide thickness, the $t_{63\%}$ of the high-k gate stack is longer.

Comparison between the results obtained by CVS and TLP methods showed that the difference between $V_{BD,TLP}$ and $V_{BD,CVS}$ is less than 10%. The TLP method can be used as an efficient method to make a conservative estimation of the breakdown voltage of the gate oxide under certain ESD events. V_{BD} data down to $t_{BD} = 1$ ns was extracted using the TLP method. The data suggests that the failure mechanism of the high-k dielectric remains the same in the 1 ns range as in the longer time scale.

For the MOSFETs in the output driver circuit, the stress is applied on the drain terminal. Only the gate dielectric in the drain-gate overlap region is under stress if the body terminal grounded. The smaller area makes the breakdown voltage larger compared to the situation when the stress is directly applied on the gate of the MOSFETs. As a result, gate dielectric breakdown in the input receiver is of more concern than in the driver. If the body of the transistor is floating, the capacitive coupling between the drain and the body will raise the body potential during an ESD event. This will increase the area of the gate dielectric under stress and hence decrease the V_{BD}.

It is demonstrated that the breakdown voltage can be increased by increasing the resistance of the current conduction path through the gate dielectric. For example, increasing the source/drain resistance or the well resistance is helpful. The well resistance can be changed through the layout of the transistor while the source/drain resistance can be increased by changing the topology of the input/output circuits. This result indicates that the immunity of the I/O circuits to GOX breakdown induced failure can be increased by the careful design of the transistor layout and the topology of the input/output circuits.

Chapter 4: Degradation of High-k/Metal Gate NMOSFETs under ESD-like Stress

4.1 Introduction

In Chapter 3, high-k gate dielectric breakdown under ESD-like stress is investigated. It was shown that for a given effective oxide thickness, the breakdown voltage V_{BD} of the high-k gate is larger than that of SiON gate. However, the presence of a large amount of electron traps, presumably oxygen vacancies [72] in the high-k layer, may result in considerable positive threshold voltage shift due to electron trapping. This has already been reported in Positive Temperature Bias Instability (PBTI) stress under low electric field conditions [13, 72, 73, 85, 86]. Electron trapping may significantly degrade transistor performance under high-field ESD-like stress. Besides, defects are created before oxide breakdown, which degrades the quality and long term reliability of the gate dielectric as well as the device performance. Several studies have been made to investigate gate oxide degradation under ESD-like stress using MOSFETs with SiO_2 or SiON gate dielectric [61, 69–71]. To ensure the reliability of high-k gate introduced in the 32 nm technology, similar studies must be made for such devices too.

In this chapter, the results of a thorough study are reported on the degradation of NMOSFETs with high-k/metal gate under non-destructive ESD-like stresses, using I-V characteristics and charge pumping measurements. The impact of stress on drain saturation current (I_{dsat}), threshold voltage (V_t), transconductance (g_m) and

subthreshold swing (SS) degradation is reported. In Sec. 4.3, we will present and discuss the effects of both positive and negative ESD-like stresses applied on the gate which emulates an ESD event on the input pin. We find that the device is more liable to degradations caused by positive stress and the degradation increases when the effective oxide thickness is increased. In addition to V_t increase and I_{dsat} decrease, the positive stress also caused considerable damage to the gate dielectric/Si interface in the NMOSFETs with thick high-k gate dielectric. Compared to SiON gate, the stress induced degradation is more severe for high-k gate. The impacts of the stress on the gate leakage current and on the PBTI degradation kinetics were also studied. In Sec. 4.4, the effects of positive stress applied on the drain are discussed, which emulates ESD events on output pins. Little degradation of the device performance was observed before device failure.

4.2 Experiment Details

The devices from the same 32 nm technology described in the last chapter were studied and the experimental setup is the same TLP system shown in Fig. 3.4. In order to evaluate the degradation caused by ESD-like stress on the input receiver, the transistor was stressed by a single 100-ns pulse on the gate with all the other terminals (source, drain and body) grounded. We monitored the impact of the pulsed stress on the performance and reliability of the NMOSFETs in both polarities ($V_G > 0$: inversion or substrate injection; $V_G < 0$: accumulation or gate injection). The stress levels were chosen within the range from ∼65% to ∼90% of the dielectric breakdown voltage (V_{BD}) corresponding to the time-to-breakdown of 100 ns. The V_{BD} data are taken from the results of the last chapter, as shown in Table 4.1.

To evaluate the degradation caused by ESD-like stress on the output driver, the

transistor was stressed on the drain terminal by a positive pulse sequence with a fixed pulse width (100 ns) and increasing amplitude. The source and the body were grounded while the gate was either grounded or positively biased. The transistor performance was monitored after each pulse until the device failed.

Table 4.1: Breakdown voltages (V_{BD}) for time-to-breakdown $t_{63\%}$ = 100 ns at 63% cumulative failure rate for SG and EG NMOSFETs in both inversion (inv) and accumulation (acc). The data were obtained by Constant Voltage Stress (CVS) measurements.

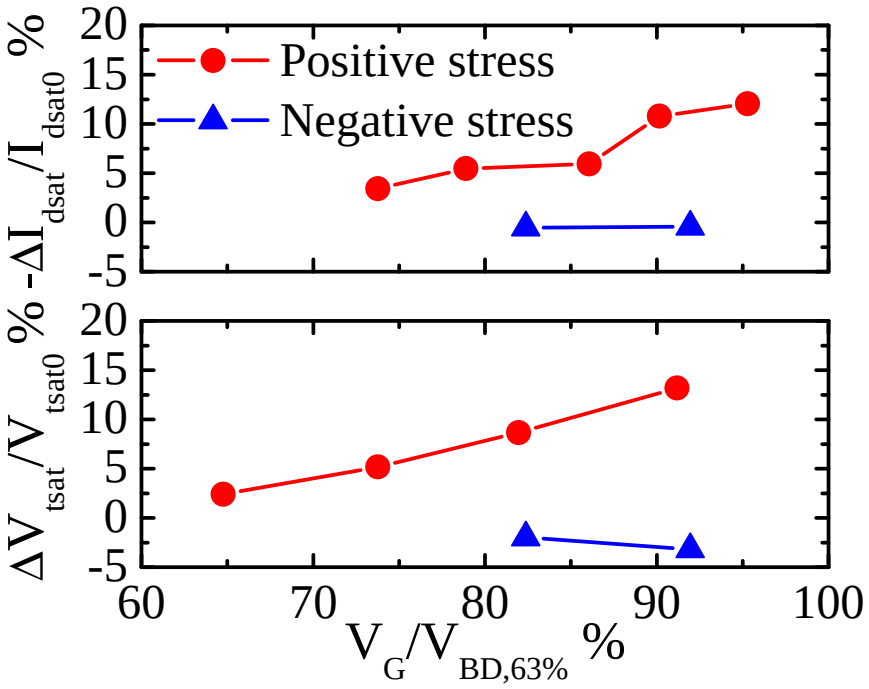
	SG Inv.	SG Acc.	EG Inv.	EG Acc.
V_{BD} (V)	4.8	5.2	9.4	12.0

4.3 Effects of ESD-like stress applied on the gate

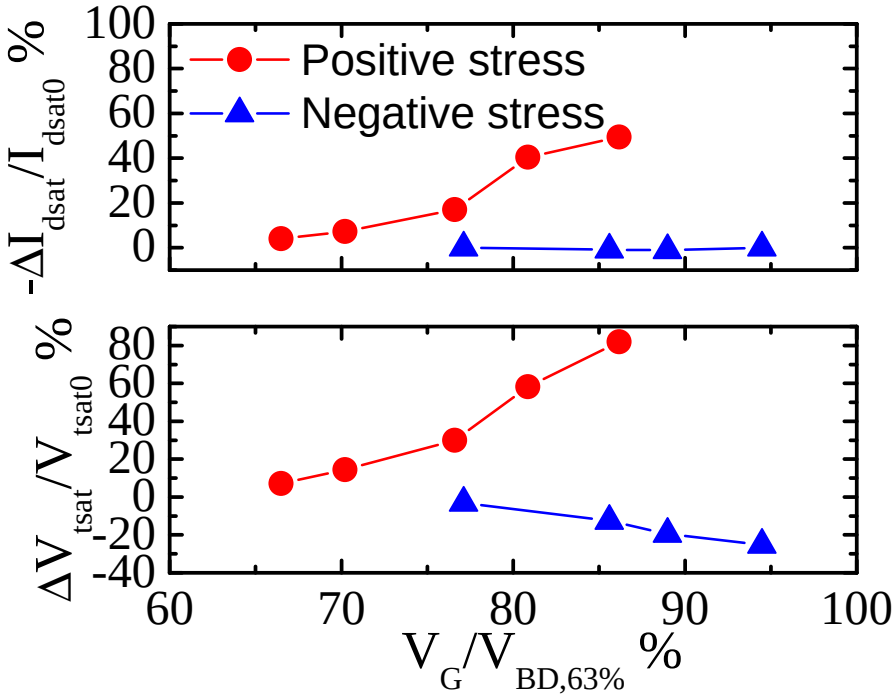
In this section, we present and discuss the effects of ESD-like stress applied on the NMOSFET gate with all the other terminals grounded. Subsection 4.3.1 considers the threshold voltage and drain current degradation. Subsection 4.3.2 considers the gate dielectric/Si interface degradation. Subsection 4.3.3 compares the degradation of high-k NMOSFETs with SiON NMOSFETs, followed by the discussion of the effects of stress on the gate leakage current in Subsection 4.3.4. Finally, Subsection 4.3.5 discusses the effect of stress on PBTI degradation kinetics.

4.3.1 Threshold voltage shift and drain current degradation

Fig. 4.1 (a) and (b) show the drain saturation current shift (ΔI_{dsat}) and the saturation threshold voltage shift (ΔV_{tsat}) at different non-destructive stress levels, for SG and EG devices respectively. The delay from stress to measurement is about 1 s for I_{dsat} and 5 s for V_{tsat}. Under positive stress, considerable increase in threshold voltage



(a) SG devices



(b) EG devices

Figure 4.1: Positive and negative pulsed stress induced degradation on drain saturation current and threshold voltage for: (a) SG and (b) EG devices. Each device was stressed by a single pulse of 100 ns long.

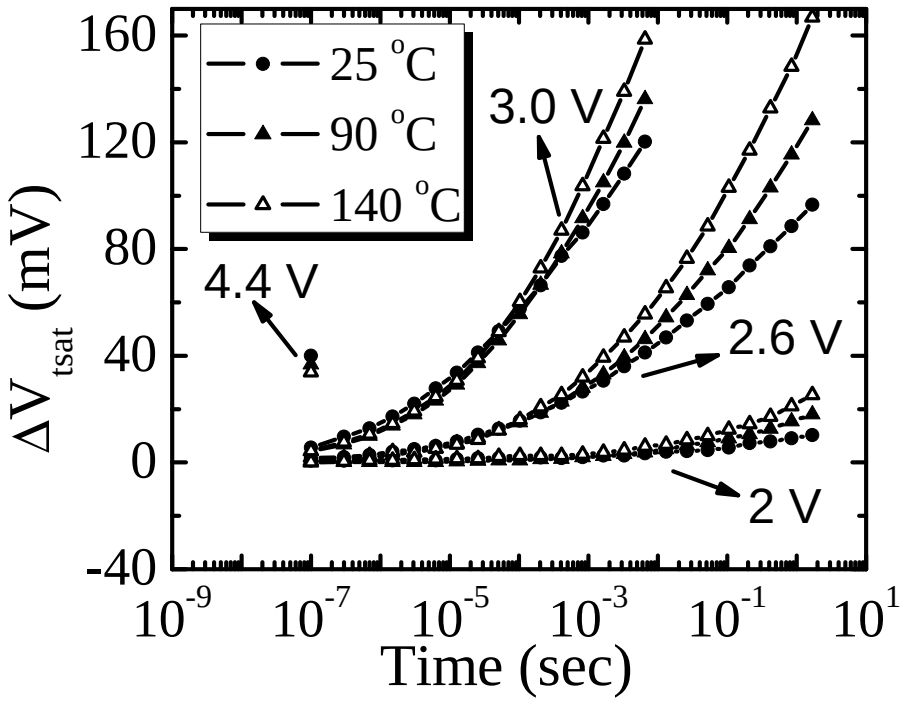


Figure 4.2: Temperature dependence of the stress induced threshold voltage shift at different stress levels as a function of cumulative stress time. A pulse sequence with constant voltage amplitude and increasing pulse width was used to stress the device gate. I_D-V_G curve was measured after each pulse to extract the threshold voltage.

and decrease in drain current were observed. Similar effects were also observed during PBTI stress, for which the stress duration is longer and the stress voltage is lower compared to pulsed stress used in this study. The V_{tsat} increase and I_{dsat} decrease can be attributed to either preexisting or stress-induced electron traps inside the gate stack [13, 72, 73, 85, 86]. Different from PBTI stress, the observed V_t shift induced by positive pulsed stress on the time scale of ESD events (∼100 ns) has a rather weak temperature dependence. Fig. 4.2 shows that for the transistor stressed at 4.4 V for 100 ns ΔV_{tsat} remains almost constant at different stress temperatures. To clarify this issue, a pulse sequence with constant voltage amplitude and an increasing pulse width was applied to stress the transistor gate. ΔV_{tsat} vs. the cumulative stress time at

different temperatures are plotted in Fig. 4.2. ΔV_{tsat} increases with temperature only when the cumulative stress time is longer than $\sim$100 μs. This suggests two charge trapping mechanisms: a fast charge trapping process with a weak temperature dependence and a slow charge trapping process with a strong temperature dependence. Similar observation was reported by Heh *et al.*[87] who proposed that the fast process is due to a resonant tunneling process and the slow process is caused by field-driven charge migration of charges trapped by the fast process.

Similar to the SiO_2 gate [61], devices with thicker high-k gate dielectric are also more liable to I_{dsat} and V_t degradations when stressed on the gate, as seen from Fig. 4.1. This can be understood by the analytical percolation model [50] which shows that the critical gate oxide defect density which causes oxide breakdown increases with the gate oxide thickness. Therefore the trapped charge density is expected to be higher in devices with thicker gate oxide, which results in more significant shifts of V_t and I_{dsat}.

The shifts of I_{dsat} and V_{tsat} immediately after a single positive pulsed stress can drive the transistor characteristics out of specifications and possibly lead to a circuit failure. If a 10% shift in V_{tsat} is taken as a failure criterion, the maximum stress level the transistor can withstand is $\sim$80% of V_{BD} for SG devices and $\sim$70% of V_{BD} for EG devices. The shift after the stress may be even larger as part of the the degradation recovers due to electron detrapping during the delay between stress and measurement limited by the experiment setup. Fig. 4.3 illustrates the transients of ΔI_{dsat} after the stress. It is shown that the recovery rate depends on the stress voltage as well as the oxide thickness. Note that I_{dsat} of SG devices recovers much faster than that of EG devices. This recovery indicates possible underestimation of ΔI_{dsat} and ΔV_{tsat} shown in Fig. 4.1, especially for SG devices. The device degradation caused by positive stress could place another limit to the upper bound of the ESD design window in

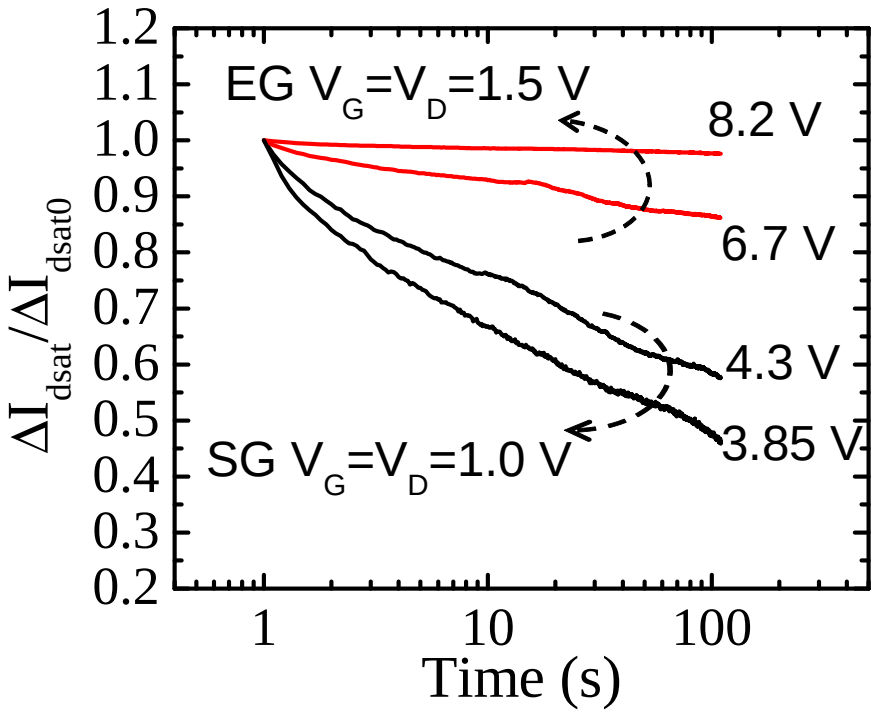


Figure 4.3: I_{dsat} transients following positive pulsed stresses at different stress levels for both SG and EG devices. The relaxation rate depends on the oxide thickness and the stress voltage.

addition to the limit resulting from the oxide breakdown voltage. This is especially so for the powered ESD events where the degradation will have an immediate impact on circuit operation before any recovery can happen.

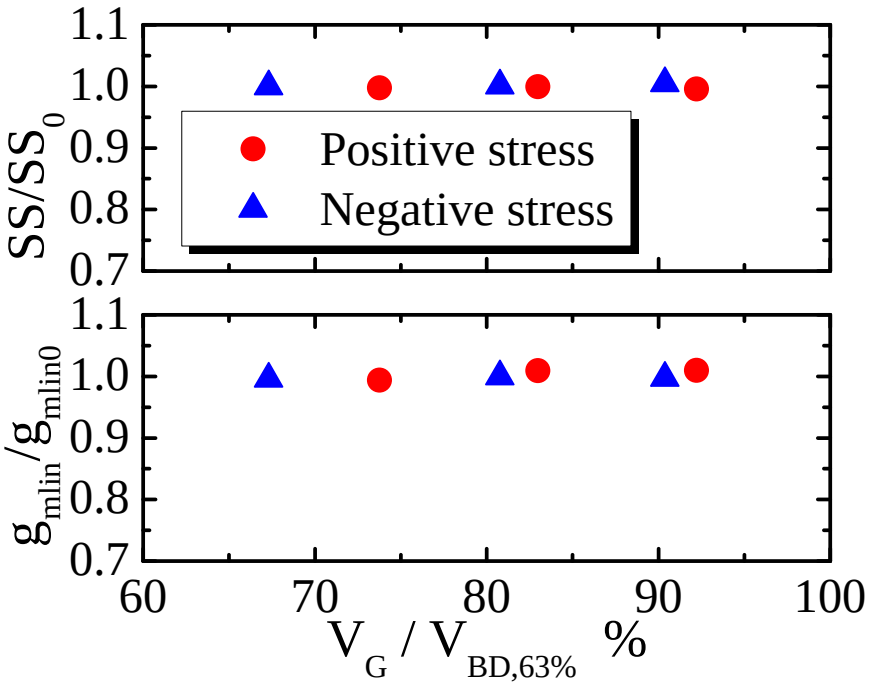
Different from positive stress which affects both SG and EG devices significantly, Fig. 4.1 shows that negative stress on the gate has noticeable effects only on the EG devices and its impact on I_{dsat} and V_{tsat} is also much smaller. Under negative ESD-like stress, the threshold voltage of the NMOSFETs is reduced, indicating positive charge trapping in the gate stack. The trapped positive charge is not stable, as 90% of the V_t shift recovers if the device is biased at 2 V for only 10 sec after the stress. Although the negative stress reduces the threshold voltage, the saturation current does not change

much because of the concurrent mobility degradation indicated by the decrease of the transconductance peak which will be discussed in the next subsection.

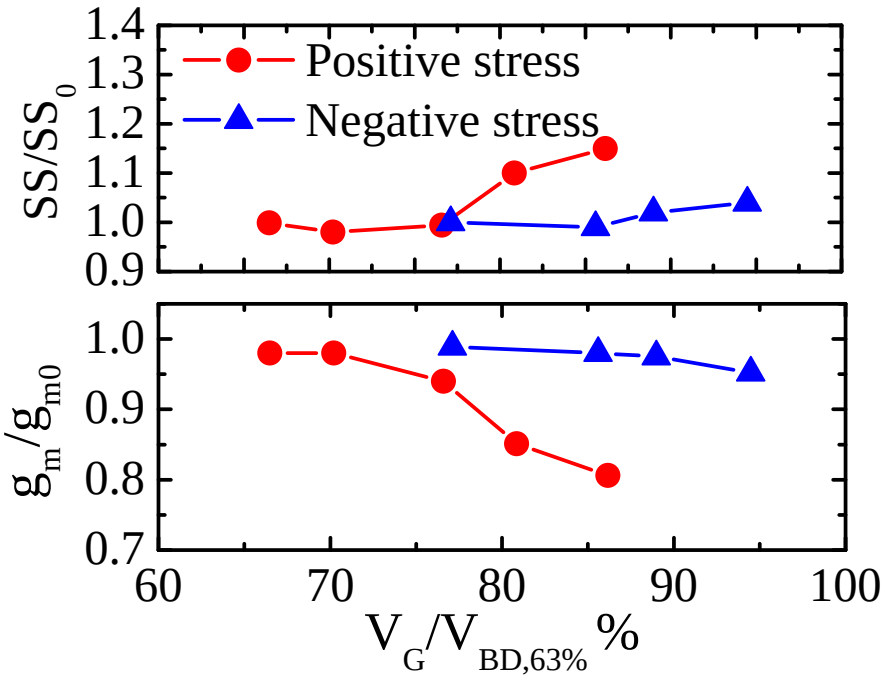
4.3.2 Degradation of the gate dielectric/Si interface

The degradation of transconductance peak (g_m) and subthreshold slope (SS) typically originates from the damage sustained by the gate dielectric/Si interface. The ESD-like stress induced interface damage was first studied by monitoring the degradation of these two parameters (extracted from the device transfer characteristics at $V_D = 0.1$ V). The results are shown in Fig. 4.4. Positive ESD-like high-voltage stress does not affect g_m and SS of the SG devices for stress levels up to ∼90% of V_{BD}. This is similar to the effect of low voltage PBTI stress (1.8V for SG and 3.0 V for EG devices up to 10^4 s) during which very little g_m and SS degradation were observed. It is also in agreement with published results, which report that for Hf, Zr and Al based high-k gate dielectrics, PBTI stress causes charge trapping or trap generation only in the bulk of the gate stack without much interface damage [73, 85, 88–90]. However, as seen in Fig. 4.4, considerable degradation of SS and g_m was observed on the EG devices at high stress levels, indicating considerable degradation sustained by the gate dielectric/Si interface. Similar to ESD-like positive stress, the degradation of SS and g_m under negative stress were also observed on the EG devices, albeit much smaller compared to that under positive stress.

The damage to the gate dielectric/Si interface of EG devices was further confirmed by using the bi-level Charge Pumping (CP) measurement [91] to measure the resulting increase of the interface trap density (N_{it}) after the transistor was stressed. The results are shown in Fig. 4.5. The trapezoidal CP wave had a constant pulse amplitude of 1.5 V and 50% duty cycle; the rise and fall time were kept at 100 ns. It is shown that positive stress creates far more interface states than negative stress at the same stress



(a) SG devices



(b) EG devices

Figure 4.4: Transconductance (g_m) and subthreshold swing (SS) degradation induced by positive and negative stress for (a) SG and (b) EG devices.

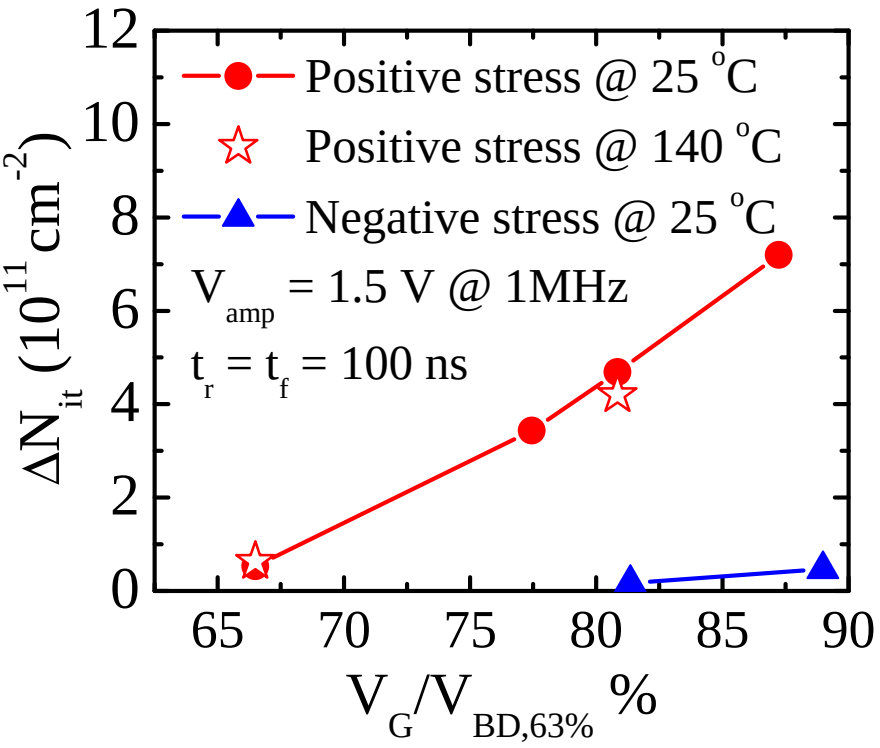


Figure 4.5: Increase of the interface state density (N_{it}) in EG devices after positive and negative stress on the gate, obtained by charge pumping measurements. Only one stress pulse was applied to each device, and the devices were stressed at either 25 °C or 140 °C. All the charge pumping measurements were made at 25 °C.

level, consistent with the more severe SS and g_m degradation under positive stress. Stressing the device at two different temperatures (25 °C and 140 °C) and performing the CP measurement at room temperature (25 °C) showed that the generation of N_{it} depends on temperature only weakly, as seen in Fig. 4.5. It has been proposed that the CP measurement with varying frequency can be used to probe traps away from Si/oxide interface [92]. For pulse frequencies scanning the range from 15 kHz to 2 MHz, the measured N_{it} changes by less than 10%, indicating most of the probed traps are indeed located at the interface.

The trivalent Si dangling bond, $Si_3{\equiv}Si\bullet$ (P_{b0} center) has been identified as one of

the most important defects at the Si/SiO_2 interface [93]. This defect is generally passivated by annealing in a hydrogen-containing atmosphere to form $Si_3{\equiv}SiH$ structure. Si-H bond breaking induced by impacting energetic carriers leads to the generation of interface states. As the energy of the carriers increases with the magnitude of the applied gate bias, interface state creation was observed only in EG devices which were stressed by a higher voltage. This is also in agreement with the result in [94] which shows that the total number of interface states (per unit area) increases with the oxide thickness. Under negative stress (gate injection), electrons are injected from the metal gate. As they tunnel through the gate oxide, these electrons gain large enough energy to break the Si-H bonds and create new interface states as they impinge on the Si/dielectric interface. For positive stress (substrate injection), the gate current is dominated by electrons injected into the gate stack at the Si/oxide interface. These cold electrons from the inversion layer typically do not have enough energy to break the Si-H bond directly. It has been suggested recently that the cold electrons can result in interface damage [95] by the reaction $Si\text{-}O+e^- \rightarrow Si^+ + O^- + e^-$ at the Si/oxide interface in devices with high-k gate. However, significant temperature dependence of the interface state generation and considerable positive charge trapping were observed in their experiments. Therefore the proposed model in [95] cannot explain the experimental results here. For $SiO_2/SiON$ gate dielectrics, the interface damage can result from two possible physical mechanisms [36, 40, 42, 96, 97]: (1) generation of holes at the anode/oxide interface (anode hole injection) or inside the gate dielectric by impact ionization; (2) build-up of hydrogen at the Si/oxide interface which is released from near the anode/oxide interface by the energetic electrons. The hydrogen may cause the depassivation of dangling silicon bonds, for example by the reaction $\equiv Si\text{-}H+H^0 \rightarrow \equiv Si\bullet + H_2$. The same mechanisms may apply to the high-k gate dielectric

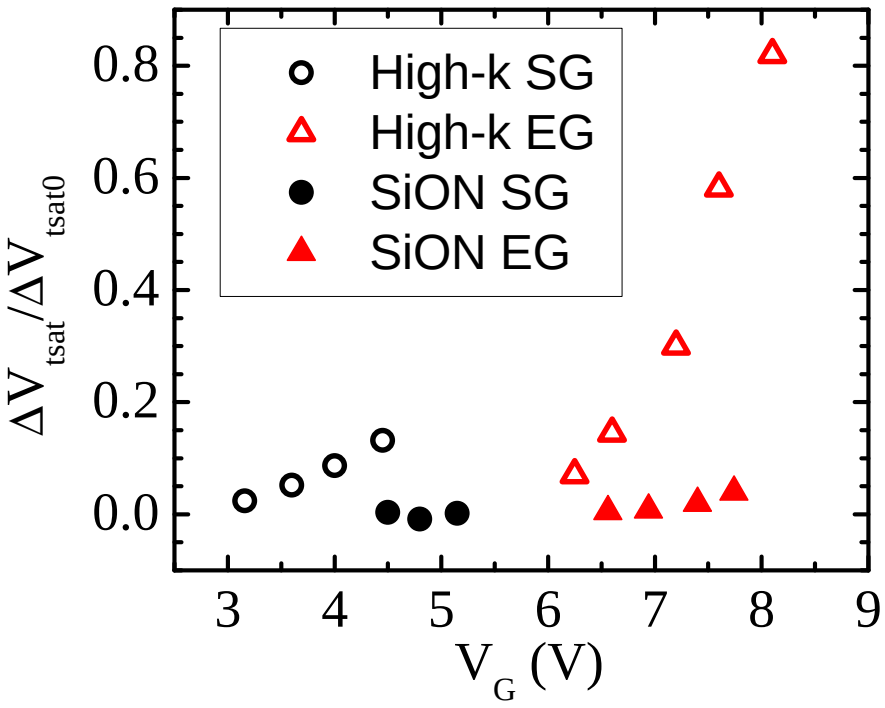


Figure 4.6: Comparison of the degradation of NMOSFETs with SiON and high-k gate under positive ESD-like stresses. More considerable V_{tsat} shift was observed in devices with high-k gate, presumably due to the large amount of electron traps inside the high-k gate dielectric.

as well. The reduced energy band gap of the high-k material (6 eV for HfO_2 compared to 9 eV for SiO_2) makes impact ionization much easier inside the gate oxide. This might be a serious reliability concern for devices with thick high-k gate dielectric under ESD-like high-field stress.

4.3.3 Comparison of SiON and high-k NMOSFETs

In this subsection, we compare the effects ESD stress on devices with SiON and high-k gate. The NMOSFETs with SiON gate are from a 45 nm bulk CMOS technology. These devices have two different effective gate oxide thicknesses: t_{inv} = 2.16 nm

(SG) and $t_{inv} = 3.4$ nm (EG). The threshold voltage shift of NMOSFETs with SiON and high-k gate dielectrics under positive stress is shown in Fig. 4.6. The stress induced V_{tsat} increase is much larger on devices with high-k gate dielectric, especially for EG devices. The degradation of I_{dsat}, g_m and SS is also much larger on high-k NMOSFETs. By contrast, these key parameter degradations induced by negative stress are similar between devices with SiON and high-k gates. Although it has been shown previously that the breakdown voltage of the high-k gate stack is larger than that of the SiO_2/SiON gate for a given effective oxide thickness [76, 98], the more serious device performance degradation sustained by the high-k NMOSFETs will certainly impact the reliability margin and draws concern on using these transistors in the input/output circuits which are directly exposed to ESD events.

4.3.4 Effect on gate leakage current

To gain more insight into the ESD-like stress induced damage on gate dielectrics, the effect of stress on gate leakage current (I_G) and its temperature dependence were studied and the results are summarized in Fig. 4.7. The gate stacks were stressed at 90% of V_{BD} for 100 ns at 25 °C and following this the gate leakage currents were measured at different temperatures in the range from 25 to 140 °C. I_G of fresh (control) devices was measured at these temperatures as well. Negative stress causes only slight increase of I_G (by less than a factor of two) for both SG and EG devices. This increase of I_G can be attributed to additional tunneling current due to trap creation in the gate stack, analogous to the stress induced leakage current (SILC) in SiO_2 gate stacks. Positive stress increases (also only slightly) the I_G of SG devices in the range of $V_G = 0.7 - 1.3$ V with a maximum at $V_G \approx 1$ V, while for voltage values smaller than 0.7 V and larger than 1.3 V I_G decreases slightly, presumably due to the trapped electrons in the gate stack which repel further electron injection [88, 89].

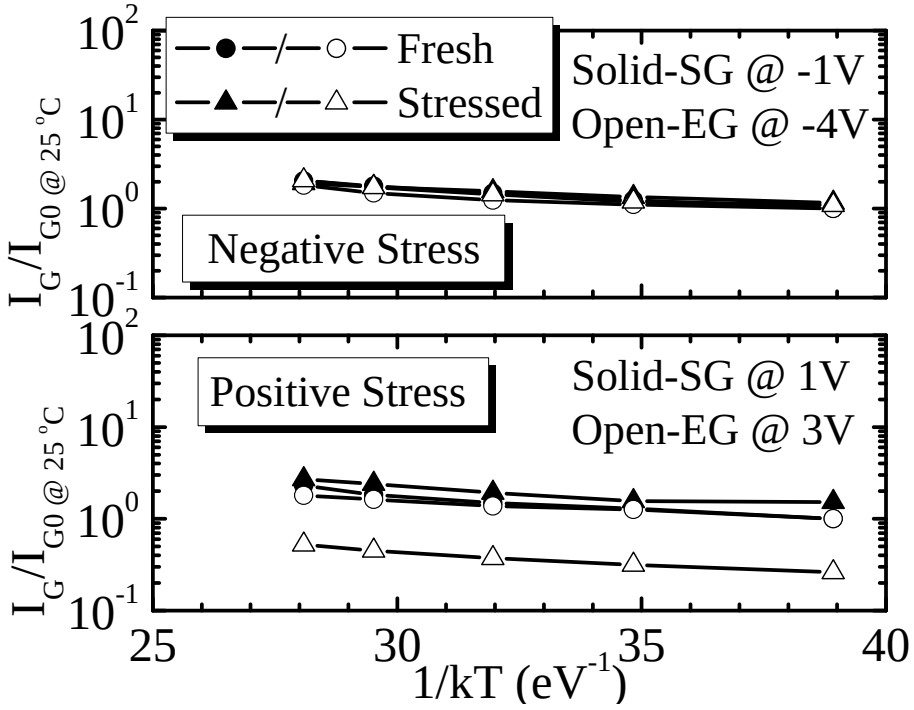


Figure 4.7: Temperature dependence of gate leakage current for fresh (control) devices and devices stress at 90% of V_{BD}. The stress was applied at 25 °C in all cases while the gate currents were measured at different temperatures.

The observed maximum at $V_G \approx 1$ V results from the alignment of the substrate Fermi level with the energy level of the created traps in the gate stack [99]. For EG devices under positive stress, I_G was observed to decrease due to significant electron trapping. Another observation from Fig. 4.7 is that the gate current has a rather weak temperature dependence both before and after stress, with activation energy of about 0.05-0.07 eV. This suggests that the generated traps lie deep inside the energy gap of the gate dielectric. The stress does not change the dominant conduction mechanism through the gate, which is either tunneling or trap-assisted tunneling through deep traps.

4.3.5 Effect on device PBTI kinetics

The large amount of carriers injected during an ESD-like stress could accelerate device degradation during subsequent electrical stress [65, 100, 101]. In order to evaluate the latent damage caused by non-destructive ESD-like stress, PBTI stress was performed on devices pre-stressed at different levels of V_{BD}. Fig. 4.8 shows the results for SG devices under positive ESD-like pre-stress. It appears that PBTI and non-destructive ESD stresses are cumulative. As some of the traps were filled during the pre-stress, the degradation rate was reduced in the initial stage of the subsequent PBTI stress. The ESD-like pre-stress does not accelerate the degradation in the long term. The same observation was made on EG devices as well. Fig. 4.9 shows the effect of negative ESD-like pre-stress on the PBTI kinetics of EG devices. Analogous to positive pre-stress, the negative pre-stress does not accelerate the PBTI degradation either. The positive charges generated in the negative pre-stress, as discussed in Sec. 4.3.1, were either detrapped or neutralized within the first 10 s of the PBTI stress and had little effect on the following PBTI stress. For the EG devices stressed under high levels of V_{BD} (86% and 94%), the pre-stress created some new traps in the gate stack, which were filled by electrons within the first 10 s of the PBTI stress and hence increased the initial ΔV_{tsat}. This shifts the $\Delta V_{tsat}/\Delta V_{tsat0}$ curve up compared to the fresh device. For SG devices, negative pre-stress up to 90% of V_{BD} has no effect on device PBTI degradation kinetics. The ΔV_{tsat} of pre-stressed and fresh devices overlapped with each other as the PBTI stress time was increased.

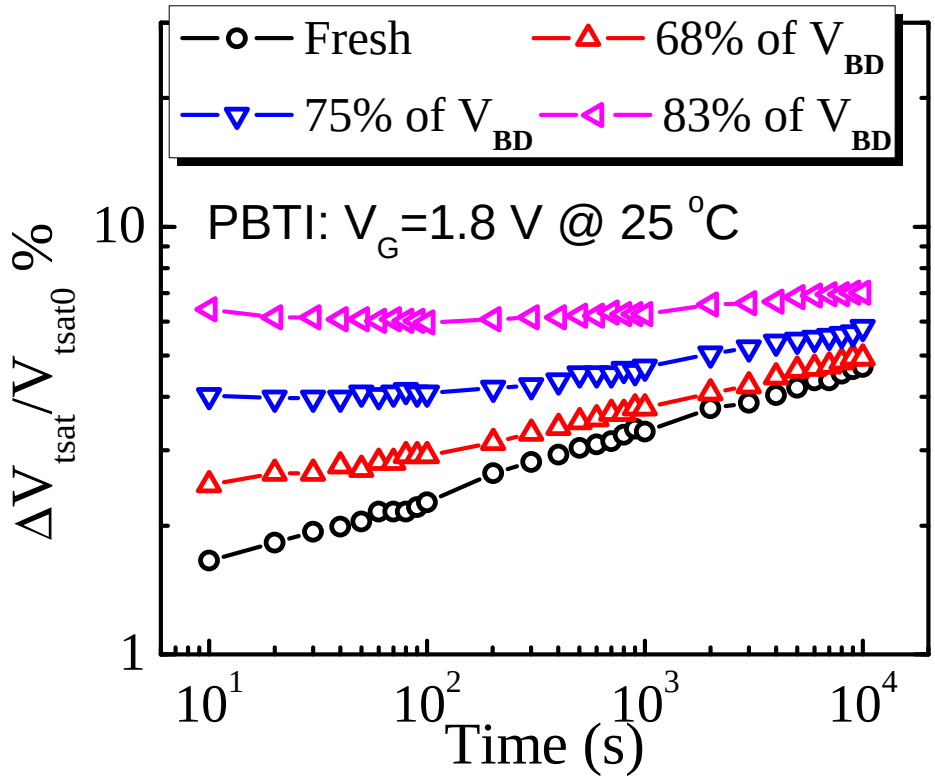


Figure 4.8: Effect of positive pre-stress on the subsequent PBTI degradation kinetics of SG devices. No pre-stress induced acceleration of the PBTI degradation kinetics was observed.

4.4 Effects of ESD-like stress applied on the drain

To emulate ESD stress on the output buffer pins, we also studied the device performance degradation induced by ESD-like stress applied on the drain of the NMOSFETs. A sequence of transmission line pulses with a constant pulse width (100 ns) and increasing amplitude was applied to the drain terminal until the transistor drain leakage current was increased by at least 10 times (taken as a failure criterion). The other terminals (gate, source and body) are kept grounded during the stress. The device linear transfer characteristics at $V_D = 0.1$ V were measured after each pulse and they are shown in Fig. 4.10. These devices failed immediately after the parasitic bipolar transistors were turned on. The SG devices failed at $V_{stress} \approx 4.8$ V and the

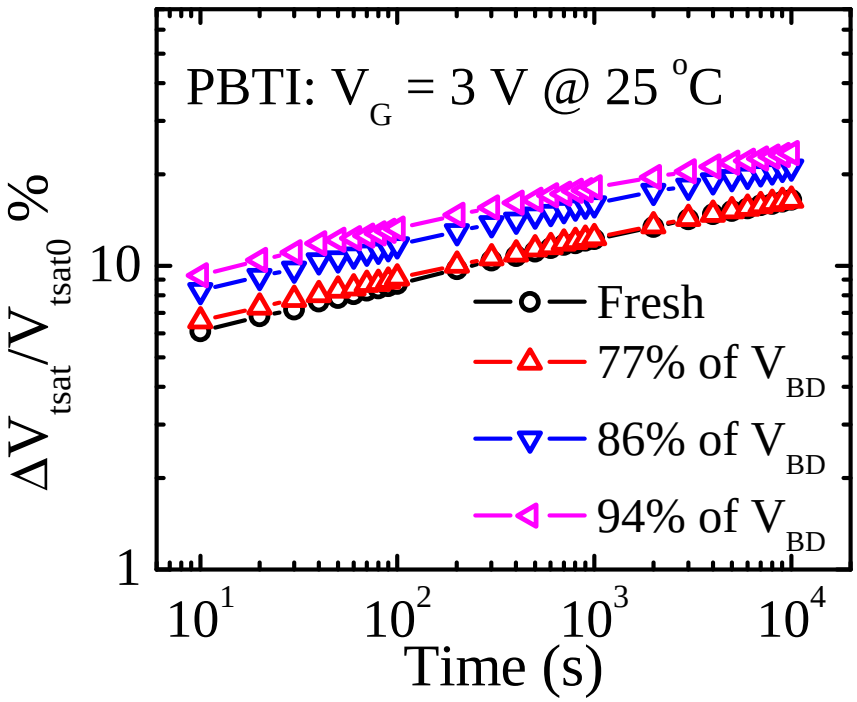


Figure 4.9: Effect of negative pre-stress on the subsequent PBTI degradation kinetics of EG devices. No pre-stress induced acceleration of the PBTI degradation kinetics was observed.

EG devices failed at $V_{stress} \approx 6.5$ V. As the source and drain of these devices were fully silicided, the parasitic bipolar transistor did not turn on uniformly [102]. The current crowding resulted in filamentation between source and drain and eventually caused device failure. As shown in Fig. 4.10, the I_D-V_G curves overlapped with each other well until the device failed and no considerable degradation was observed prior to failure. The small degradation can be attributed to the following two reasons. First, when the stress is applied on the drain, the effective gate dielectric area under stress is smaller than that when the stress is applied on the gate as the stress voltage drops along the channel and thus the breakdown voltage is expected to be higher than stressing the gate directly. Second, stressing the drain positively with

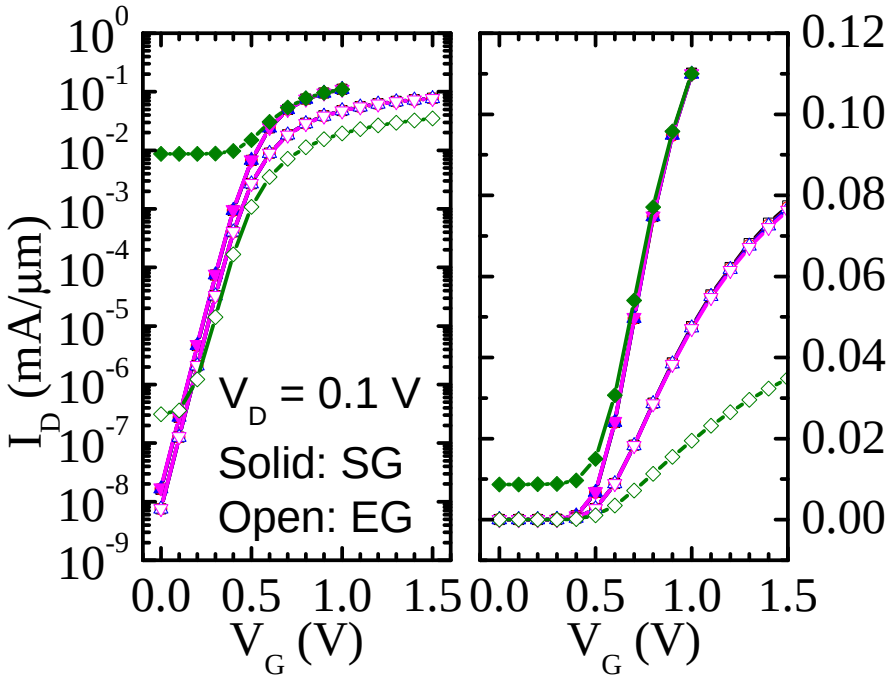


Figure 4.10: The effect of positive ESD-like stresses on the transfer characteristics of SG and EG devices in logarithmic (left) and linear (right) scales. The stress was applied on the drain with gate, source and body grounded. No significant device degradation was observed as the stress voltage was ramped up until the device failed.

gate grounded is analogous to stressing the gate negatively with drain grounded and according to the data presented in Sec. 4.3 the degradation will be small. As the gate bias may be raised during an ESD event due to capacitive coupling, experiments were also performed with a DC bias applied to the gate (1.0 V for SG and 1.5 V for EG devices). The gate bias only reduces the parasitic bipolar turn-on voltage, and no other degradation was observed until the device failed.

4.5 Summary

In this chapter, the device performance degradation induced by ESD-like stress on state of the art NMOSFETs with high-k/metal gate stacks was studied using devices

from a 32 nm bulk CMOS technology.

To emulate ESD events on the output pins, positive pulses were applied to the drain of the NMOSFETs. Little degradation in threshold voltage (V_t), drain saturation current (I_{dsat}), transconductance (g_m) and subthreshold swing (SS) was observed as the stress voltage was ramped up until the device failed. The failure is dominated by drain to source filamentation rather than oxide breakdown in these fully silicided devices.

To emulate ESD events on the input pins, positive and negative pulsed stresses were applied to the gate of the NMOSFETs. The stress changes the gate leakage current only marginally and does not accelerate the subsequent PBTI degradation kinetics. However, positive ESD-like stress decreases the I_{dsat} and increases the V_t of the NMOSFETs significantly, more so for the devices with thicker gate dielectric. This degradation can be attributed to considerable electron trapping in the high-k gate stack. By contrast to typical PBTI stress, here the V_t shift depends on temperature rather weakly, which indicates that a different dominant charge trapping mechanism is active on the time scale of ESD events. In addition to V_t and I_{dsat} degradation, positive ESD-like stress on the EG devices also causes significant damage to the Si/oxide interface, which degrades g_m and SS. Furthermore, positive stress caused much more degradation on transistors with high-k gate than with SiON gate. Finally, it was observed that the impact of negative stress on I_{dsat}, V_t, g_m and SS is much smaller than that of positive stress. All these observed key parameter degradations induced by positive ESD-like stress will affect the device reliability and impact the upper bound of the ESD design window.

Chapter 5: Design and Optimization of the SOI Field Effect Diode (FED) for ESD Protection

5.1 Introduction

The scaling of semiconductor technology unfavorably impacts on-chip ESD protection performance by decreasing the I/O MOSFET's second breakdown [1] current density (I_{t2}) and oxide breakdown voltage (V_{BD}) [103]. The decline of current shunting capability of the ESD devices exacerbates the design challenge. The reduction of metal interconnect thickness also degrades the current carrying capability. All these trends imply the shrinkage of the ESD design window. Moreover, with the more stringent capacitance budget for high-speed I/O it is important to find ESD solutions that minimize the parasitic loading effects while achieving superior robustness. In addition to the above-mentioned challenges in ESD protection for advanced CMOS technologies, Silicon-on-Insulator (SOI) technology has distinctive challenges because of the buried oxide (BOX) layer between the active device and the silicon substrate.

Silicon-on-insulator (SOI) has gained wide acceptance for future technology nodes to keep the scaling trend due to performance superiority over similar bulk technologies [103]. However, the unavailability of vertical current and the self-heating effect in SOI CMOS because of the presence of the buried oxide layer (BOX) makes simple integration of common bulk ESD protection devices more difficult. Despite these, several SOI ESD protection devices, such as lateral diodes and silicide-blocked grounded

[1] Second breakdown [27] occurs at a higher current injection level than avalanche breakdown. Second breakdown causes permanent damage to the device and the onset of second breakdown is defined as the damage threshold of the device.

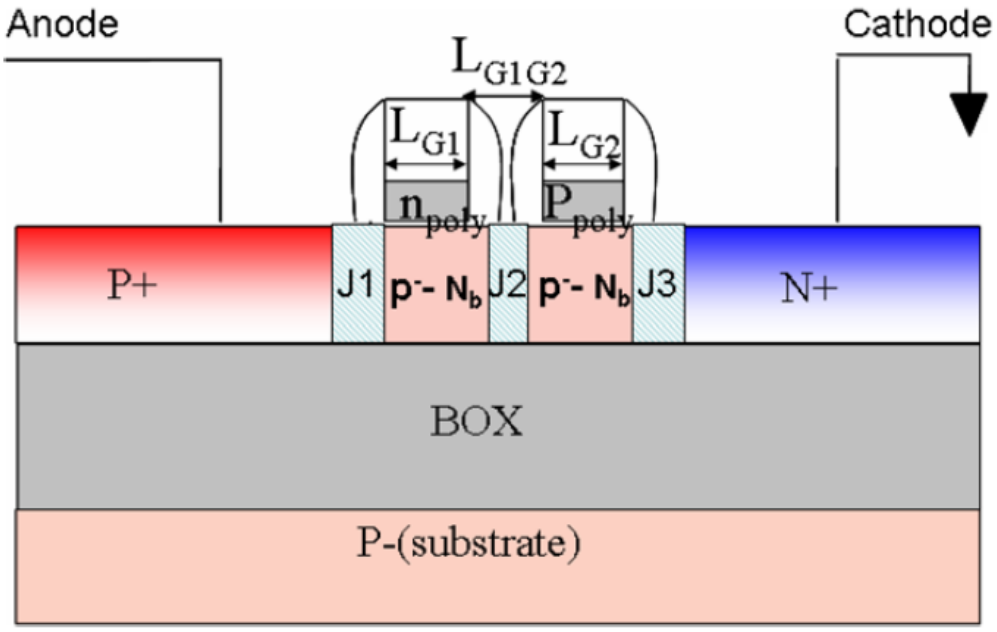


Figure 5.1: Cross-section of the two-gate FED with gates denoted as G1 and G2. J2 is induced by positive V_{G1}. It is clear that the structure resembles a pin diode, augmented by two gates between the anode and the cathode.

gate NMOSFET, have been proposed. However, these devices either do not allow local clamping or have relatively low second-breakdown current. Consequently, the ESD protection for SOI devices remains a challenging issue. A new SOI device, the so called Field Effect Diode (FED), has been recently reported which has interesting applications in ESD protection [21]. The importance of this structure lies in the fact that by appropriately biasing it, its operation can be switched between semiconductor controlled rectifier (SCR)-like and diode-like. Fig. 5.1 shows the cross-section of a standard FED, where it is seen that it resembles a lateral SOI diode, augmented by two closely spaced gates: gate 1 (G1) on the anode (p+) side and gate 2 (G2) on the cathode (n+) side. By biasing G1 positively and G2 negatively (or grounded) the channel region beneath G1 gets inverted and a temporary junction J2 is created. The device acts like an SCR in this state, while grounding G1 and G2 makes the device behave as a normal diode. Fig. 5.2 shows typical I-V characteristics of the device,

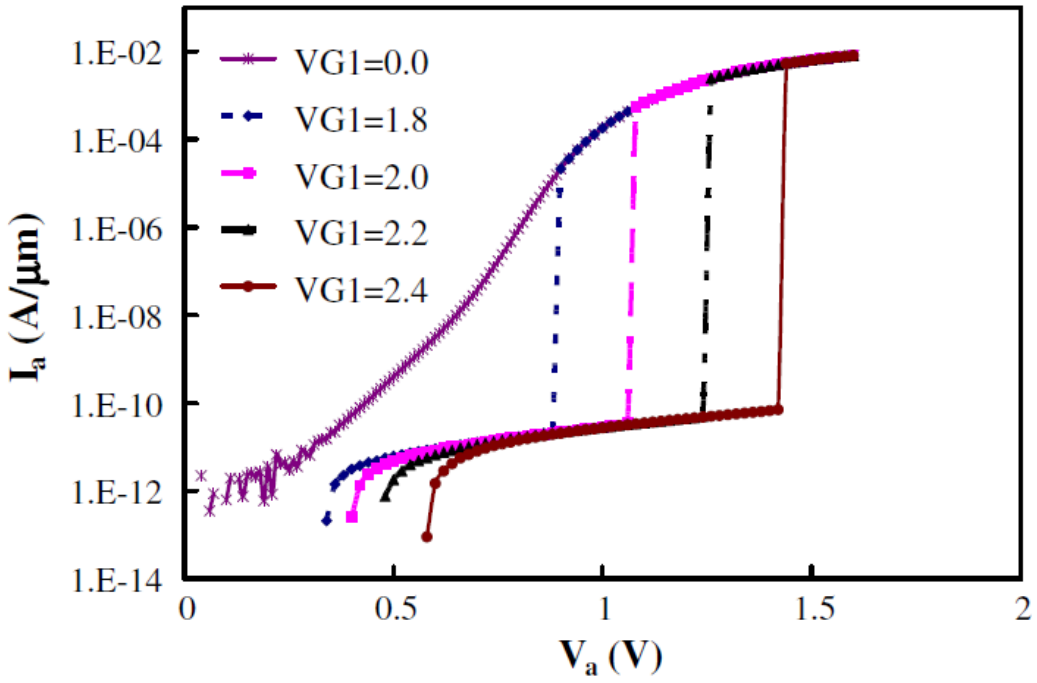


Figure 5.2: Measured DC ICV curves for the two-gate FED ($N_b = 6e16\text{cm}^{-3}$ p-type, $T_{soi} = 70$ nm, $L_{G1} = L_{G2} = 0.5\mu$m) with varying gate voltage V_{G1} ($V_{G2} = 0$ V).

where it is seen that the breakdown voltage can be controlled by different gate-biasing schemes.

An example of the application of FED in ESD protection is shown in Fig. 5.3 where a regular pn diode is used for negative and a FED for positive ESD pulses on the Input/Output (I/O) pad . The two gates of the FED are biased to achieve a high anode breakdown voltage V_{FB} so that the device is off in normal situations, while the two gates are grounded during positive ESD pulses. This local I/O clamping scheme is very important because the discharge path through resistive busing and a supply clamp is eliminated. A large V_{FB} is needed to keep the device off in normal situations. To date these required large V_{FB} values have only been achieved by applying large gate voltages, incompatible with the current low power CMOS technology (1 V typically). This chapter is therefore intended to explore the device design space in an effort to

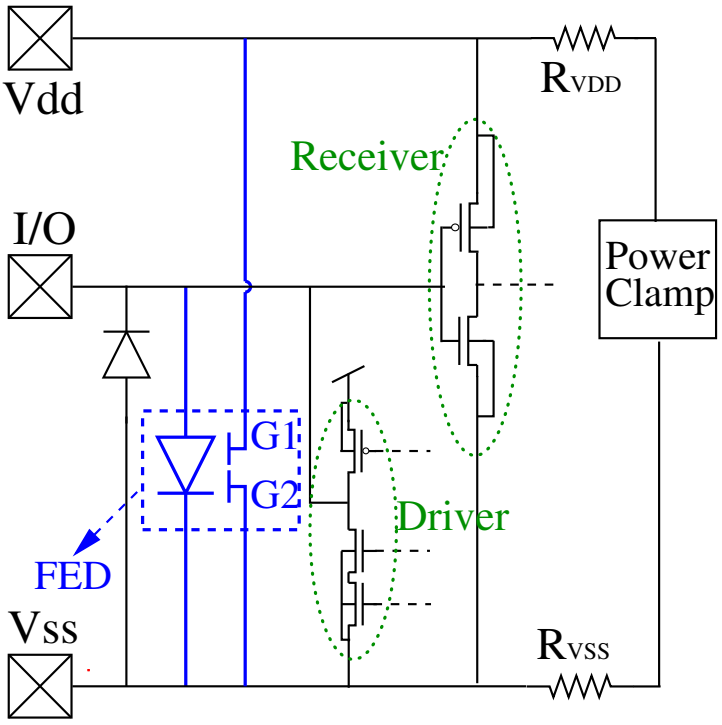


Figure 5.3: Schematic of an I/O circuit using a regular lateral diode for negative ESD pulse protection and FED for positive ESD pulse protection (a local clamping scheme).

see if a large V_{FB} can be achieved without the application of excessively large gate voltages.

5.2 Simulation of FED

FED design and optimization are investigated using 2D isothermal simulations. Mobility models [104,105] are used to account for the effect of impurity, surface roughness and carrier-carrier scattering. Shockley-Read-Hall carrier lifetimes are assumed to be the same for both electrons and holes for simplicity and their doping dependence is accounted for based on the Scharffeter relation. The carrier lifetime mentioned in the next section is the maximum carrier lifetime used in the doping-dependence model.

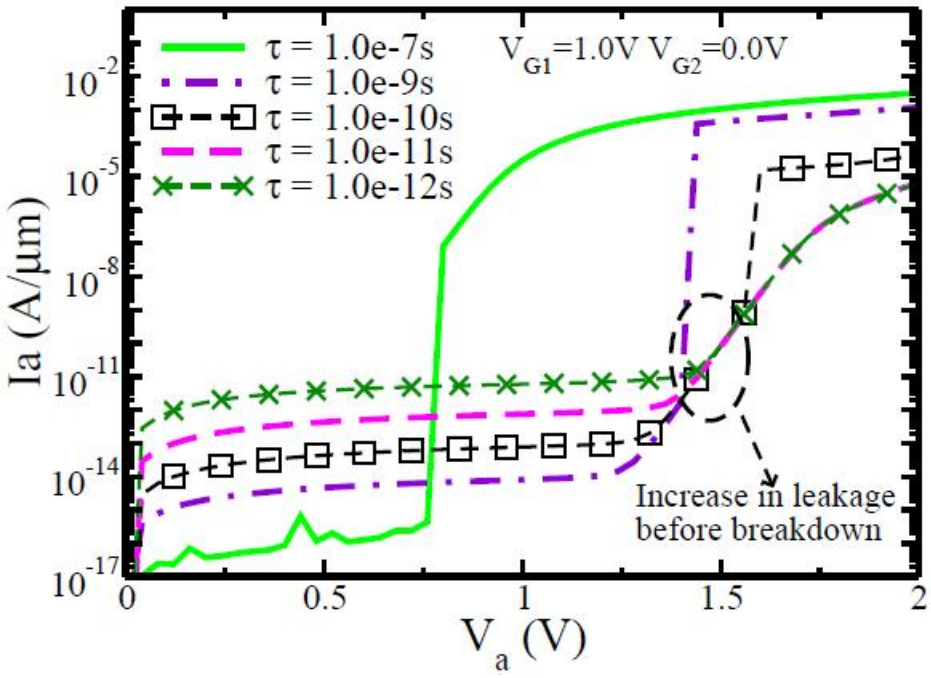


Figure 5.4: Effect of carrier lifetime on the simulated I-V curves of the "standard" two gate FED.

To improve simulation speed and stability in the presence of breakdown, transient bias sweep is used for both DC and transient simulations. SOI layer thickness is set to be 70 nm in all simulations. Body doping concentration N_b and its type, doping concentration N_{G2} beneath G2, and gate lengths L_{G1} and L_{G2} are all varied in a systematic way, and the results are compared to the "standard" device with $L_{G1} = L_{G2} = 0.5$ μm, $N_b = 1e16$ cm^{-3} p-type, $N_{gap} = 0$ and $N_{G2} = 0$.

5.3 FED Design Space and Discussion

The breakdown voltage can be expressed as $V_{FB} = V_B(1-\alpha_1-\alpha_2)$ [106], where V_B is the breakdown voltage of the PN junction (J2) between the two gates and α_1 and α_2 are the dc common base gains of the NPN and PNP bipolar transistors intrinsic to

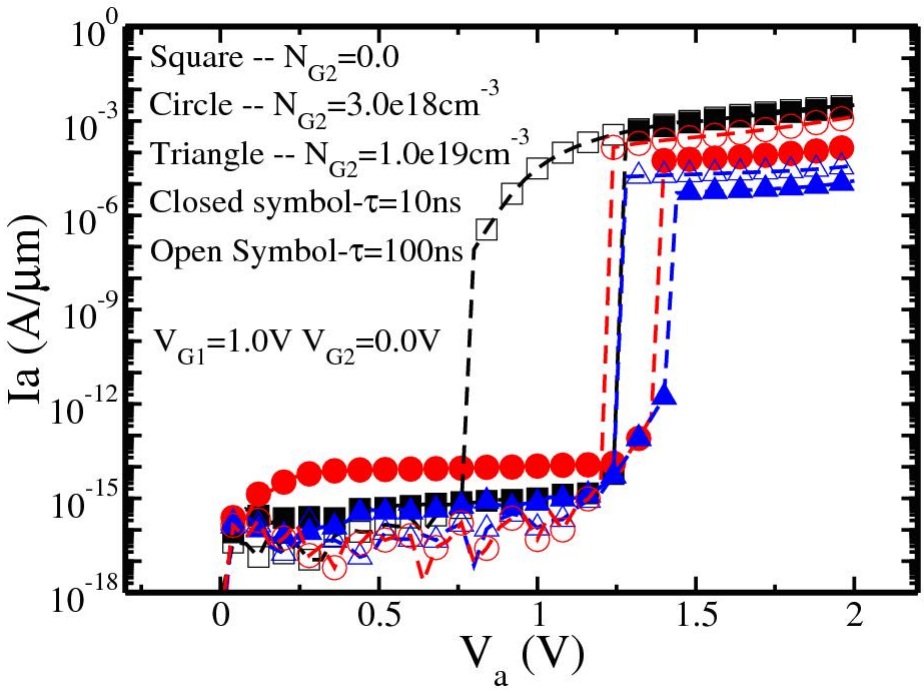
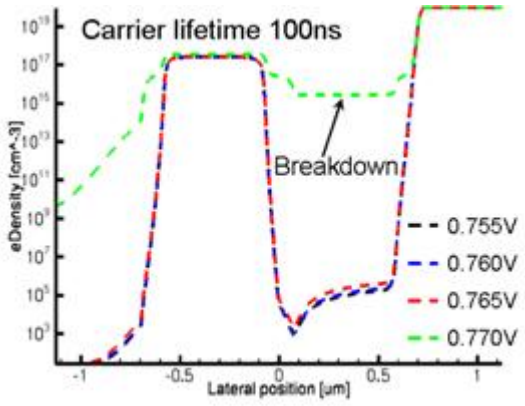
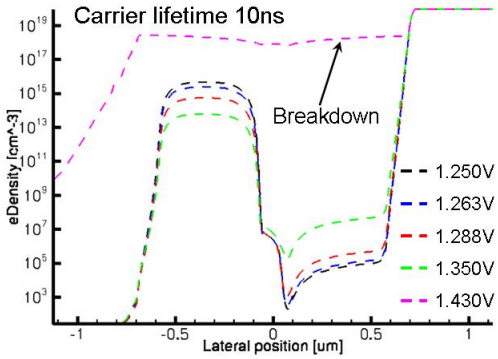


Figure 5.5: Effect of doping concentration N_{G2} beneath G2 on the I-V curves of the FED.

the FED structure. Because α_1 and α_2 are strongly dependent on the carrier lifetime, V_{FB} also strongly depends on the carrier lifetime and is actually expected to increase by reducing the lifetime. However, the device in the on state can be treated as a p-i-n diode, of which the current driving capability degrades significantly with decreasing carrier lifetime [107]. These effects are shown in Fig. 5.4. Fig. 5.5 shows that V_{FB} can also be increased by increasing the p-type doping concentration under G2, which decreases the gain α_2 of the bipolar transistor. However, both Fig. 5.4 and Fig. 5.5 show that V_{FB} saturates to a certain level independent of the carrier lifetime and the base doping (base Gummel number) of the NPN bipolar transistor. Increasing the anode bias raises the electron quasi Fermi level beneath G1. As a result, the negative charges in the MOS capacitor formed by G1 and the SOI layer are mainly inversion electrons at SiO_2/Si interface at large anode bias, while most of the SOI layer is left



(a) Carrier lifetime $\tau = 100$ ns



(b) Carrier lifetime $\tau = 10$ ns

Figure 5.6: Simulated electron densities 0.01μm below the Si/SiO_2 interface in the "standard" FED with different carrier lifetimes. Voltage ranges from 0.755 V to 0.77 V in (a) and from 1.25 V to 1.43 V in (b). As shown in (b), electron density decreases significantly near breakdown.

unaffected by the gate bias. This is shown in Fig. 5.6 where only a few electrons exist beneath G1 at large anode voltage Va. This eliminates the induced junction J2 and causes breakdown of the device. The depletion of electrons beneath G1 thus imposes an upper limit for V_{FB} for this standard device. Small carrier lifetime or large p-type doping concentration beneath G1 should be used to achieve sufficiently large V_{FB}. However, as mentioned above these methods increase both leakage current and on-resistance of the device.

To push V_{FB} beyond this limit, the region beneath G1 should be designed so

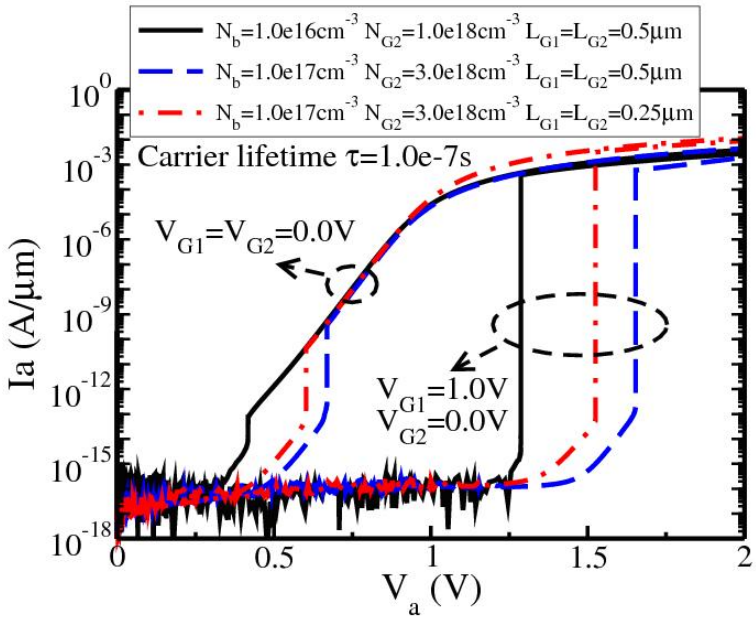


Figure 5.7: Simulated I-V curves for the two-gate FED with uniform n-type body doping under G1.

that electron density remains high at large V_a. This can be achieved through n-type doping beneath G1 and p-type doping beneath G2, to form a PNPN structure with two gates G1 and G2. Fig. 5.7 shows the IV characteristics, where it is seen that in this case V_{G1} modulates V_{FB} over a wide range. Similarly to the FED case where the SOI film is doped p-type, increasing the anode bias tends to invert the PMOSFET formed by the p-type anode/G1/p-type region beneath G2, which leads to device breakdown. This assists hole injection into the p-type region beneath G1 and hence the regenerative process of the SCR structure. High n-type doping concentration beneath G1 is therefore preferable for a larger breakdown voltage as it increases the threshold voltage of the PMOSFET. The surface potential and gate voltage of a fully

depleted (FD) SOI MOSFET are related by the following equations [108]:

$$V_{G1} = \Phi_{ms1} + \left(1 + \frac{C_{si}}{C_{ox1}}\right)\Phi_{s1} - \frac{C_{si}}{C_{ox1}}\Phi_{s2} - \frac{Q_{depl}}{2C_{ox1}}, \tag{5.1}$$

$$V_{G2} = \Phi_{ms2} - \frac{C_{si}}{C_{ox2}}\Phi_{s1} + \left(1 + \frac{C_{si}}{C_{ox2}}\right)\Phi_{s2} - \frac{Q_{depl}}{2C_{ox2}}, \tag{5.2}$$

where it is assumed that the inversion charge density is zero. Subscripts 1 and 2 indicate the front gate oxide and buried oxide respectively. For devices with relatively higher doping concentration (which means small carrier lifetime), the breakdown voltage V_{FB} can be estimated by assuming weak inversion condition under G1. From the above equations the V_{FB} of the device with $N_b = 1e17$ cm^{-3} (n-type) is calculated to be around 1.6 V, in agreement with the simulated value in Fig. 5.7. However, for the device with lower doping concentration, i.e. $N_b = 1e16$ cm^{-3}, the breakdown happens before G1 is weakly inverted due to its relatively larger carrier lifetime (the sum of the common dc transistor gain α_1 and α_2 of the PNP and NPN transistors approaches 1 before G1 is weakly inverted).

As the device no longer behaves as a diode when $V_{G1} = 0.0$ V, there might be a concern with regard to the switching speed of the device. In bulk technologies, the time needed to charge the base-emitter junction capacitance of the NPN and PNP devices (the main component of the switching speed) is minimized by increasing the well resistance [109]. Here this is achieved inherently as a result of the floating bases of both PNP and NPN transistors and the reduced junction capacitance inherent in SOI technology. As shown in Fig. 5.7, L_{G1} and L_{G2} can be reduced to a certain level without compromising V_{FB} very much; this is helpful to reduce the base transit time and hence increase the turn-on speed of the device. Fig. 5.8 compares the transient behavior of the modified devices with the "standard" device under a 3.9 mA current

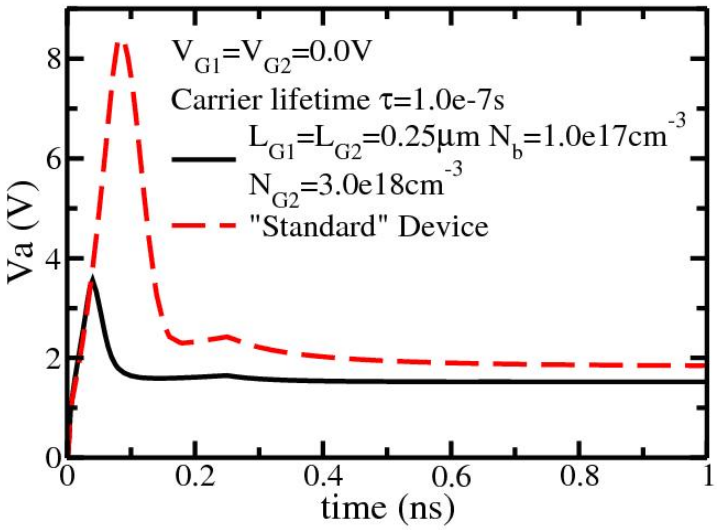


Figure 5.8: Transient behavior of different device structures subjected to current pulses.

pulse with rise time of 250 ps; it shows comparable switching speed and a much lower voltage peak during the transient. Shorter gate lengths are also beneficial for lowering on-resistance (Fig. 5.7) and for area-efficiency.

5.4 Experimental verification

The device structure proposed here based on numerical simulation was later proved to be useful by Cao et. al. [110] based on the experimental results. They showed that V_{FB} can be increased by controlling the doping concentration under G1 and G2 separately. For example, increasing the p-type doping concentration under G2 and doping the region under G1 with n-type dopant are both helpful to increase the anode breakdown voltage V_{FB}. Fig. 5.9 [110] shows that $V_{G1} = 1$ V is large enough to make V_{FB} larger than 1 V (which is the nominal power supply voltage for the latest technology node), with increased p-type doping concentration under G1. Compared to the I-V curve of the original FED shown in Fig. 5.2, the anode breakdown voltage

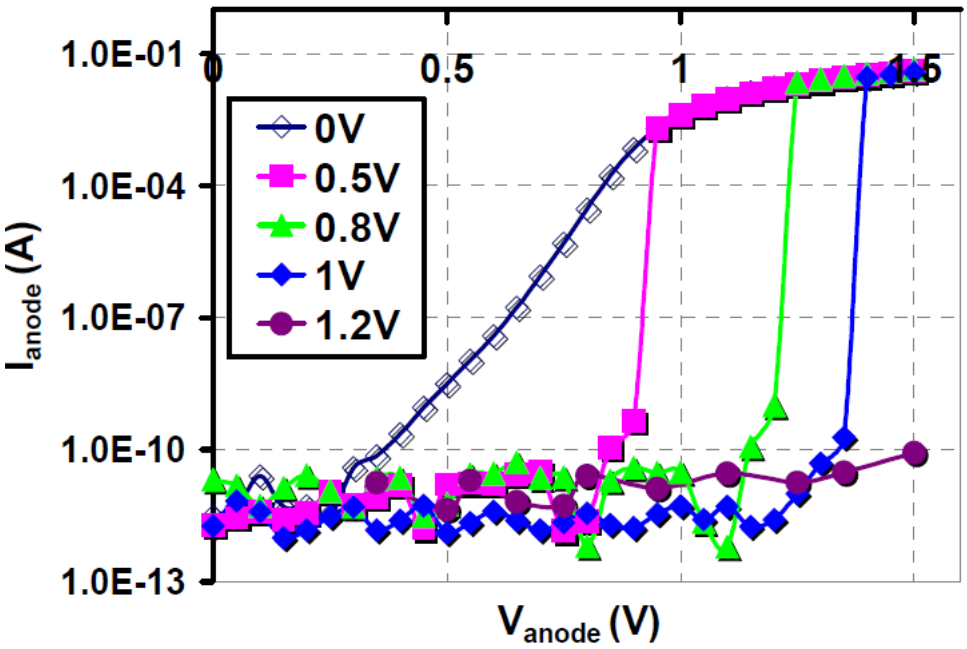


Figure 5.9: Measured DC I-V of FED with increased p-type doping concentration under G2, with different G1 biasing.

V_{FB} is increased considerably.

5.5 Summary

For FED devices with p-type body doping, increasing the p-type doping concentration under G2 is helpful to increase the anode breakdown voltage V_{FB} as high doping concentration decreases the bipolar gain and hence increases the bipolar turn-on voltage. However, increasing the anode voltage suppresses the depletion layer width beneath G1. This keeps the SOI layer unaffected by the G1 gate voltage, due to the raised electron quasi Fermi potential of the p-type SOI layer. This imposes an upper limit to the anode breakdown voltage V_{FB}. Numerical simulations have demonstrated that the use of n-type doping beneath G1 and p-type doping beneath G2 (PNPN structure) leads to sufficiently high breakdown voltages to avoid unintended device latchup.

The simulations also showed that switching speed of the proposed PNPN device is comparable to the "standard" FED because of the nature of the SOI structure. The study also demonstrates that higher turn-on speed and lower on-resistance can be achieved by reducing the channel length. The study by [110] provides experimental evidence that the proposed methods are helpful to achieve V_{FB} high enough for real applications.

Chapter 6: Conclusions and future work

6.1 Conclusions

In this dissertation, we have studied ESD related issues in nano-scale CMOS technologies. The main focus has been on understanding high-k gate dielectric breakdown and degradation under ESD like stress first. The most important conclusions of this work can be summarized as follows:

- It is confirmed that high-k gate dielectric breakdown is catastrophic under ESD-like pulsed stress. The high voltage stress changes the gate leakage current only slightly until the final hard breakdown. After gate oxide failure, the large gate current may result in a loss of gate contact, and short the drain and the source terminals.

- As the entire gate is stressed, the transistors in the input receiver are more susceptible to gate dielectric breakdown compared to the transistors in the output driver. For the stress configuration of the input receiver, the results of NMOSFETs and PMOSFETs under both positive and negative stress show that NMOSFETs under positive stress have the smallest V_{BD}.

- Comparison between the breakdown voltage obtained by CVS ($V_{BD,CVS}$) and TLP methods ($V_{BD,TLP}$) shows that the difference between $V_{BD,TLP}$ and $V_{BD,CVS}$ is less than 10%. The TLP method can thus be used as an efficient method to make a conservative estimation of the breakdown voltage of the gate oxide under different ESD events.

- V_{BD} data down to $t_{BD} = 1$ ns were also extracted using the TLP method. These 1-ns data are close to the data extrapolated from the results in the 50 ns to 10 μs time domain, indicating that the failure mechanism of the high-k dielectric remains the same in the 1 ns range as in the longer time scale.

- It is demonstrated that the breakdown voltage can be increased by increasing the resistance of the current conduction path through the gate dielectric. For example, increasing the source/drain resistance or the well resistance is helpful. The well resistance can be changed through the layout of the transistor while the source/drain resistance can be increased by changing the topology of the input/output circuits. This result indicates that the immunity to gate dielectric breakdown can be improved by the careful design of the transistor layout and the topology of the input/output circuits.

- The dissertation then shows that device performance degradation is considerable for transistors in the input receiver, where the gate terminal is directly exposed to ESD events. The degradation is more significant under positive stress, possibly due to the large amount electron traps inside the high-k gate. After the positive stress, threshold voltage (V_t) is increased and drain saturation current (I_{dsat}) is decreased. The degradation increases with the effective oxide thickness. By contrast to typical PBTI stress, the V_t shift under ESD-like stress depends on temperature rather weakly, which implies that a different dominant charge trapping mechanism is active on the time scale of ESD events. Different from typical PBTI stress, considerable interface states are generated after positive ESD-like stress in the thick oxide device, which degrades both the transconductance g_m and the subthreshold slope SS.

- The comparison between MOSFETs with high-k gate and SiON gate showed

that for a given effective gate oxide thickness, the breakdown voltage of the high-k gate stack is larger. However, transistors with high-k gate are more susceptible to device performance degradation under positive ESD-like stress.

The dissertation also conducted a thorough investigation of the field effect diode (FED) with the aim to explore its potential for ESD protection applications in silicon on insulator (SOI) technology. Based on the TCAD simulation results, this work shows that by increasing the p-type doping concentration under gate 2 or doping the region under gate 1 with n-type dopant, FED devices with reasonable breakdown voltage values can be achieved at the gate biasing compatible with the latest technology. The proposed methodology is verified experimentally by the other research group [110].

6.2 Future work

In this dissertation, gate dielectric breakdown voltage corresponding to the cumulative failure rate of 63% is discussed. This value can not be used directly as the maximum allowable voltage the gate dielectric can tolerate, as 63% of the measured samples have already failed. Breakdown voltage at lower failure percentile requires accurate knowledge of the slope of the Weibull distribution. However, as shown in Fig. 3.8, the slope of the cumulative distribution is not a constant. This bimodal distribution has also been reported by other researchers [75, 111, 112]. Further understanding the statistics of the high-k gate breakdown in the ESD-time domain will be an useful extension of this work.

This study also shows that the TLP method can be used as a conservative estimation of the breakdown voltage. It will be of great interest to find out if the TLP meausrement results follow the area scaling law of gate dielectric breakdown.

As discussed in Chapter 3, the gate dielectric is not always stressed on the gate terminal with other terminals grounded. For example, capacitive coupling or series resistance effect may change the voltage at the device terminals. For SOI devices, the potential of the transistor body might be changed due to charges introduced by the tunneling current. Therefore, a reliable model to predict gate dielectric breakdown in these different stress conditions will be useful. Besides, the study in [62] pointed out that hard breakdown results in a unique state for the transistor. To enable fast post-breakdown circuit simulation for determining the logic function failure, these authors also developed a macromodel of post-breakdown transistor for 90 nm and 130 nm technologies [63, 64]. A extension of this model to transistors with high-k gate dielectric will be an interesting and important topic to study.

The results in Chapter 4 show that significant damage to the Si/gate oxide interface is caused by positive stress. The degradation in threshold voltage and drain saturation current is also considerable. This undesirable behavior may be improved by optimizing the process flow. However, a deeper understanding of the physical origin of the degradation will be certainly required. Chapter 4 also shows that the ESD-press does not accelerate the PBTI degradation kinetics. However, the pre-stress indeed increases the initial threshold voltage shift. This may affect the device lifetime. Further study is required to clarify this issue.

CDM-ESD events have a very short discharging time and extremely large discharging current. The typical peak current is 10 A with a rise time of only $\sim$ 200 ps. Consequently, the turn-on speed of the protection device will be a key issue. CDM-ESD events will be the major threat to transistor gate dielectric in the future technologies. The data related to sub-ns gate dielectric breakdown and degradation in this study is very limited. A more detailed study in this short time domain is required.

The extreme scaling of transistor dimensions is expected to end for CMOS bulk technologies in the near future. Simply shrinking the dimensions is not sufficient to guarantee performance any more. New materials and novel device structures are required. Because of the superior control over the short-channel-effects and good scaling possibility, fully-depleted SOI MOSFETs with ultrathin body (UTB) and FinFET devices [108] have been proposed to be the promising candidates to continue the CMOS scaling trend beyond the 22 nm technology node. Meanwhile, novel channel materials, such as Ge and III-V ($In_{0.7}Ga_{0.3}As$) [113, 114], are investigate to enhance channel mobility. ESD performance of these new devices will be an important issue to be addressed in the future.

Appendix A: Relationship between $V_{BD,CVS}$ and $V_{BD,TLP}$

Based on the power law relationship between $t_{63\%}$ and V_{BD}, stressing the device at V_{G1} for time t_1 is equivalent to stressing the device at V_{G2} for time $t_2 = t_1 \cdot (V_{G1}/V_{G2})^n$. Assuming the duration of the ESD event is t_{pw}, the breakdown voltage, $V_{BD,CVS}$, corresponding to t_{pw} (obtained from CVS method) can be related to the result from the TLP measurement with the pulse width equal to t_{pw} by Eq. (A.1) [53]:

$$t_{pw} = \sum_{i}^{m} t_{pw} \cdot \left(\frac{V_i}{V_{BD,CVS}} \right)^n . \tag{A.1}$$

V_i is the TLP voltage in the i^{th} step, m is the number of voltage pulses until the gate oxide failure and n is the power law exponent. The breakdown voltage from the TLP method is $V_{BD,TLP}$, corresponding to the magnitude of the last voltage pulse. $V_{BD,TLP}$ is equal to $m \cdot \Delta V$. Approximating the summation in Eq. (A.1) by an integral yields

$$t_{pw} = \int_0^{m \cdot t_{pw}} \left(\frac{RR \cdot t}{V_{BD,CVS}} \right)^n dt. \tag{A.2}$$

$RR = \Delta V / t_{pw}$ and ΔV is the voltage increment in each TLP step. Calculating this integration gives

$$t_{pw} = \frac{V_{BD,CVS}}{RR \cdot (n+1)} \left(\frac{V_{BD,TLP}}{V_{BD,CVS}} \right)^{n+1} . \tag{A.3}$$

Reformulating Eq. (A.3) gives the relationship between $V_{BD,TLP}/V_{BD,CVS}$ and $V_{BD,CVS}$:

$$\frac{V_{BD,TLP}}{V_{BD,CVS}} = \left(\frac{\Delta V \cdot (n+1)}{V_{BD,CVS}}\right)^{1/(n+1)} . \tag{A.4}$$

Fig. 3.19 is obtained by plotting $V_{BD,TLP}/V_{BD,CVS}$ against n and $V_{BD,CVS}$.

Bibliography

Bibliography

[1] H. Wong and H. Iwai, "On the scaling issues and high-k replacement of ultrathin gate dielectrics for nanoscale MOS transistors," *Microelectronic Engineering*, vol. 83, no. 10, pp. 1867 – 1904, 2006.

[2] Y. Taur and T. H. Ning, *Fundamentals of modern VLSI devices, 2nd ed.* Cambridge University Press, 2009.

[3] J. McPherson, J. Kim, A. Shanware, H. Mogul, and J. Rodriguez, "Trends in the ultimate breakdown strength of high dielectric-constant materials," *Electron Devices, IEEE Transactions on*, vol. 50, no. 8, pp. 1771–1778, Aug. 2003.

[4] [Online]. Available: http://en.wikipedia.org/wiki/Transistor_count

[5] C. Duvvury, "Paradigm shift in esd qualification," in *Reliability Physics Symposium, IEEE International*, 2008, pp. 1 –2.

[6] S. Beebe, "Characterization, modeling, and design of esd protection circuits," Ph.D. dissertation, Stanford University, 1998.

[7] O. Semenov, H. Sarbishaei, and M. Sachdev, *ESD Protection Device and Circuit Design for Advanced CMOS Technologies.* Springer, 2008.

[8] "International technology roadmap for semiconductors, 2007 edition," Tech. Rep.

[9] E. Wu and R.-P. Vollertsen, "On the Weibull shape factor of intrinsic breakdown of dielectric films and its accurate experimental determination. Part I: theory, methodology, experimental techniques," *Electron Devices, IEEE Transactions on*, vol. 49, no. 12, pp. 2131–2140, Dec 2002.

[10] E. Miranda and J. Sune, "Analytic modeling of leakage current through multiple breakdown paths in SiO_2 films," in *Reliability Physics Symposium, IEEE International*, 2001, pp. 367–379.

[11] J. Sune, E. Wu, and W. Lai, "Statistics of competing post-breakdown failure modes in ultrathin MOS devices," *Electron Devices, IEEE Transactions on*, vol. 53, no. 2, pp. 224 – 234, 2006.

[12] T. W. Chen, C. Ito, T. Maloney, W. Loh, and R. Dutton, "Gate oxide reliability characterization in the 100ps regime with ultra-fast transmission line pulsing system," in *EOS/ESD Symp.*, 2007.

[13] A. Kerber and E. Cartier, "Reliability challenges for CMOS technology qualifications with Hafnium Oxide/Titanium Nitride gate stacks," *Device and Materials Reliability, IEEE Transactions on*, vol. 9, no. 2, pp. 147–162, June 2009.

[14] J. Lilienfeld, "Method and apparatus for controlling electric currents," U.S. Patent 1 745 175, 1930.

[15] D. Kahng and M. M. Atalla, "Silicon-silicon dioxide field induced surface devices," in *IRE Solid-State Device Res. Conf.*, 1960.

[16] W. Haensch, E. J. Nowak, R. H. Dennard, P. M. Solomon, A. Bryant, O. H. Dokumaci, A. Kumar, X. Wang, J. B. Johnson, and M. V. Fischetti, "Silicon cmos devices beyond scaling," *IBM Journal of Research and Development*, vol. 50, no. 4, pp. 339 –361, July 2006.

[17] J. Plummer and P. Griffin, "Material and process limits in silicon vlsi technology," *Proceedings of the IEEE*, vol. 89, no. 3, pp. 240 –258, mar 2001.

[18] R. H. DENNARD, F. H. GAENSSLEN, H.-N. YU, V. LEO RIDEOVT, E. BASSOUS, and A. R. LEBLANC, "Design of ion-implanted mosfet with very small physical dimensions," *Solid-State Circuits Newsletter, IEEE*, vol. 9, pp. 256–268, 1974.

[19] G. Baccarani, M. Wordeman, and R. Dennard, "Generalized scaling theory and its application to a 1/4 micrometer mosfet design," *Electron Devices, IEEE Transactions on*, vol. 31, no. 4, pp. 452 – 462, apr 1984.

[20] S. Voldman, R. Schulz, J. Howard, V. Gross, S. Wu, A. Yapsir, D. Sadana, H. Hovel, J. Walker, F. Assaderaghi, B. Chen, J.-C. Sun, and G. Shahidi, "CMOS-on-SOI ESD protection networks," in *EOS/ESD Symp.*, 10-12 1996, pp. 291 – 301.

[21] A. Salman, S. Beebe, M. Emam, M. Pelella, and D. Ioannou, "Field effect diode (fed): A novel device for esd protection in deep sub-micron soi technologies," in *Electron Devices Meeting, IEEE International*, dec. 2006, pp. 109–112.

[22] T. Green, "A review of eos/esd field failures in military equipment," in *EOS/ESD Symp.*, 1988, pp. 7–14.

[23] R. G. Wagner, J. Soden, and C. F. Hawkins, "Extent and cost of EOS/ESD damage in an IC manufacturing process," in *EOS/ESD Symp.*, 1993, pp. 49–55.

[24] L. van Roozendaal, A. Amerasekera, P. Bos, W. Baelde, F. Bontekoe, P. Kersten, E. E. Korma, P. Rommers, P. Krys, U. Weber, and P. Ashby, "Standard esd testing of integrated circuits," in *EOS/ESD Symp.*, 1990, pp. 119–130.

[25] K. Verhaege, P. J. Roussel, G. Groeseneker, H. E. Maes, H. Gieser, C. Russ, P. Egger, X. Guggenmos, and F. G. Kuper, "Analysis of hbm esd testers and specifications using a 4th order lumped element model," in *EOS/ESD Symp.*, 1993, pp. 129–137.

[26] T. Maloney and N. Khurana, "Transmission line pulsing techniques for circuit modeling of ESD phenomena," in *EOS/ESD Symp.*, 1985, pp. 49–54.

[27] A. Amerasekera and C. Duvvury, *ESD in Silicon integrated circuits.* John Wiley & Sons, Ltd., 2002.

[28] N. Klein, "The mechanism of self-healing electrical breakdown in mos structures," *Electron Devices, IEEE Transactions on*, vol. 13, no. 11, pp. 788–805, Nov 1966.

[29] M. Shatzkes, M. Av-Ron, and R. M. Anderson, "On the nature of conduction and switching in SiO_2," *Journal of Applied Physics*, vol. 45, no. 5, pp. 2065–2077, 1974.

[30] J. W. McPherson and D. A. Baglee, "Acceleration factors for thin gate oxide stressing," in *Proc. Int. Reliab. Phys. Symp.*, March 1985, pp. 1–5.

[31] J. McPherson, "Stress dependent activation energy," in *Reliability Physics Symposium, IEEE International*, 1986, pp. 12 –18.

[32] J. W. McPherson and H. C. Mogul, "Disturbed bonding states in SiO_2 thin-films and their impact on time-dependent dielectric breakdown," in *Proc. Int. Reliab. Phys. Symp.*, 31 March–2 April 1998, pp. 47–56.

[33] J. W. McPherson, R. B. Khamankar, and A. Shanware, "Complementary model for intrinsic time-dependent dielectric breakdown in sio2 dielectrics," *Journal of Applied Physics*, vol. 88, no. 9, pp. 5351 –5359, 2000.

[34] J. W. McPherson, "Determination of the nature of molecular bonding in silica from time-dependent dielectric breakdown data," *Journal of Applied Physics*, vol. 95, no. 12, pp. 8101 –8109, 2004.

[35] D. J. DiMaria, E. Cartier, and D. Arnold, "Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon," *Journal of Applied Physics*, vol. 73, no. 7, pp. 3367 –3384, 1993.

[36] K. Schuegraf and C. Hu, "Hole injection SiO_2 breakdown model for very low voltage lifetime extrapolation," *Electron Devices, IEEE Transactions on*, vol. 41, no. 5, pp. 761 –767, May 1994.

[37] D. J. DiMaria and J. W. Stasiak, "Trap creation in silicon dioxide produced by hot electrons," *Journal of Applied Physics*, vol. 65, no. 6, pp. 2342 –2356, 1989.

[38] D. J. DiMaria, "Defect generation under substrate-hot-electron injection into ultrathin silicon dioxide layers," *Journal of Applied Physics*, vol. 86, no. 4, pp. 2100–2109, 1999.

[39] M. Alam, J. Bude, and A. Ghetti, "Field acceleration for oxide breakdown-can an accurate anode hole injection model resolve the E vs. 1/E controversy?" in *Proc. Int. Reliab. Phys. Symp.*, 2000, pp. 21–26.

[40] J. Sune and E. Wu, "Hydrogen-release mechanisms in the breakdown of thin SiO_2 films," *Phys. Rev. Lett.*, vol. 92, no. 8, pp. 0 876 011–0 876 014, 2004.

[41] J. Sune and E. Y. Wu, "Mechanisms of hydrogen release in the breakdown of SiO_2-based gate oxides," in *Electron Devices Meeting, IEEE International*, 5–5 Dec. 2005, pp. 388–391.

[42] P. E. Nicollian, A. T. Krishnan, C. A. Chancellor, R. B. Khamankar, S. Chakravarthi, C. Bowen, and V. K. Reddy, "The current understanding of the trap generation mechanisms that lead to the power law model for gate dielectric breakdown," in *Reliability Physics Symposium, IEEE International*, 15–19 April 2007, pp. 197–208.

[43] E. M. Vogel, J. S. Suehle, M. D. Edelstein, B. Wang, Y. Chen, and J. B. Bernstein, "Reliability of ultrathin silicon dioxide under combined substrate hot-electron and constant voltage tunneling stress," *Electron Devices, IEEE Transactions on*, vol. 47, no. 6, pp. 1183–1191, June 2000.

[44] L. Selmi, D. Esseni, and P. Palestri, "Towards microscopic understanding of MOSFET reliability: the role of carrier energy and transport simulations," in *Electron Devices Meeting, IEEE International*, 2003, pp. 333–336.

[45] E. Wu, J. Sune, and W. Lai, "On the Weibull shape factor of intrinsic breakdown of dielectric films and its accurate experimental determination. Part II: experimental results and the effects of stress conditions," *Electron Devices, IEEE Transactions on*, vol. 49, no. 12, pp. 2141–2150, Dec 2002.

[46] J. Su, I. Placencia, N. Barniol, E. Farrs, F. Martin, and X. Aymerich, "On the breakdown statistics of very thin SiO_2 films," *Thin Solid Films*, vol. 185, pp. 347–362, 1990.

[47] D. J. Dumin, S. K. Mopuri, S. Vanchinathan, R. S. Scott, R. Subramoniam, and T. G. Lewis, "High field related thin oxide wearout and breakdown," *Electron Devices, IEEE Transactions on*, vol. 42, no. 4, pp. 760–772, April 1995.

[48] R. Degraeve, G. Groeseneken, R. Bellens, J. L. Ogier, M. Depas, P. J. Roussel, and H. E. Maes, "New insights in the relation between electron trap generation and the statistical properties of oxide breakdown," *Electron Devices, IEEE Transactions on*, vol. 45, no. 4, pp. 904–911, April 1998.

[49] J. H. Stathis, "Percolation models for gate oxide breakdown," *Journal of Applied Physics*, vol. 96, no. 10, pp. 5757–5766, 99.

[50] J. Sune, "New physics-based analytic approach to the thin-oxide breakdown statistics," *Electron Device Letters, IEEE*, vol. 22, no. 6, pp. 296–298, Jun 2001.

[51] B. E. Weir, C.-C. Leung, P. J. Silverman, and M. A. Alam, "Gate dielectric breakdown in the time-scale of ESD events," *Microelectronics Reliability*, vol. 45, no. 3-4, pp. 427–436, 2005.

[52] A. Ille, W. Stadler, A. Kerber, T. Pompl, T. Brodbeck, K. Esmark, and A. Bravaix, "Ultra-thin gate oxide reliability in the ESD time domain," in *EOS/ESD Symp.*, 2006, pp. 185–194.

[53] A. Kerber, L. Pantisano, A. Veloso, G. Groeseneken, and M. Kerber, "Reliability screening of high-k dielectrics based on voltage ramp stress," *Microelectronics Reliability*, vol. 47, no. 4-5, pp. 513 – 517, 2007.

[54] K. P. Cheung, "Impact of ESD protection device trigger transient on the reliability of ultra-thin gate oxide," *Microelectronics Reliability*, vol. 41, no. 5, pp. 745–749, 2001.

[55] S. Lombardo, J. H. Stathis, B. P. Linder, K. L. Pey, F. Palumbo, and C. H. Tung, "Dielectric breakdown mechanisms in gate oxides," *Journal of Applied Physics*, vol. 98, no. 12, p. 121301, 2005.

[56] K. Matsuzawa, H. Satake, C. Sutou, and H. Kawashima, "Simulation of number of pulses to breakdown during TLP for ESD testing," in *SISPAD*, Sept. 2003, pp. 129–132.

[57] J. Wu, P. Juliano, and E. Rosenbaum, "Breakdown and latent damage of ultra-thin gate oxides under ESD stress conditions," in *EOS/ESD Symp.*, 2000, pp. 287–293.

[58] J. Wu and E. Rosenbaum, "Gate oxide reliability under ESD-like pulse stress," *Electron Devices, IEEE Transactions on*, vol. 51, no. 9, pp. 1528–1532, Sept. 2004.

[59] B. E. Weir, C.-C. Leung, P. J. Silverman, and M. A. Alam, "Gate dielectric breakdown: A focus on ESD protection," in *Reliability Physics Symposium, IEEE International*, 2004, p. 399404.

[60] E. Wu, A. Vayshenker, E. Nowak, J. Sune, R.-P. Vollertsen, W. Lai, and D. Harmon, "Experimental evidence of TBD power-law for voltage dependence of oxide breakdown in ultrathin gate oxides," *Electron Devices, IEEE Transactions on*, vol. 49, no. 12, pp. 2244–2253, Dec 2002.

[61] A. Ille, W. Stadler, T. Pompl, H. Gossner, T. Brodbeck, K. Esmark, P. Riess, D. Alvarez, K. Chatty, R. Gauthier, and A. Bravaix, "Reliability aspects of gate oxide under ESD pulse stress," in *EOS/ESD Symp.*, 2007, pp. 328–337.

[62] T. Chen, C. Ito, W. Loh, and R. Dutton, "Post-breakdown leakage resistance and its dependence on device area," *Microelectronics Reliability*, vol. 46, pp. 1612–1616, 2006.

[63] T. W. Chen, C. Ito, W. Loh, W. Wang, S. Mitra, and R. Dutton, "Macromodel for post-breakdown 90nm and 130nm transistors and its applications in predicting chip-level function failure after ESD-CDM events," in *Reliability Physics Symposium, IEEE International*, 2007, pp. 78 –85.

[64] T. W. Chen, C. Ito, W. Loh, W. Wang, K. Doddapaneni, S. Mitra, and R. Dutton, "Design methodology and protection strategy for esd-cdm robust digital system design in 90-nm and 130-nm technologies," *Electron Devices, IEEE Transactions on*, vol. 56, no. 2, pp. 275 –283, feb. 2009.

[65] S. Aur, A. Chatterjee, and T. Polgreen, "Hot-electron reliability and ESD latent damage," *Electron Devices, IEEE Transactions on*, vol. 35, no. 12, pp. 2189–2193, Dec 1988.

[66] B. Doyle, D. Krakauer, and K. Mistry, "Examination of oxide damage during high-current stress of n-MOS transistors," *Electron Devices, IEEE Transactions on*, vol. 40, no. 5, pp. 980–985, May 1993.

[67] A. Salman, R. Gauthier, W. Stadler, K. Esmark, M. Muhammad, C. Putnam, and D. Ioannou, "Characterization and investigation of the interaction between hot electron and electrostatic discharge stresses using NMOS devices in 0.13 μm CMOS technology," in *Reliability Physics Symposium, IEEE International*, 2001, pp. 219–225.

[68] A. Salman, R. Gauthier, C. Putnam, P. Riess, M. Muhammad, M. Woo, and D. Ioannou, "ESD-induced oxide breakdown on self-protecting GG-nMOSFET in 0.1-μm CMOS technology," *Device and Materials Reliability, IEEE Transactions on*, vol. 3, no. 3, pp. 79–84, Sept. 2003.

[69] A. Cester, S. Gerardin, A. Tazzoli, and G. Meneghesso, "Electrostatic discharge effects in ultrathin gate oxide MOSFETs," *Device and Materials Reliability, IEEE Transactions on*, vol. 6, no. 1, pp. 87–94, March 2006.

[70] J.-C. Tseng and J.-G. Hwu, "Effects of electrostatic discharge high-field current impulse on oxide breakdown," *Journal of Applied Physics*, vol. 101, no. 1, p. 014103, 2007.

[71] ——, "Oxide-trapped charges induced by electrostatic discharge impulse stress," *Electron Devices, IEEE Transactions on*, vol. 54, no. 7, pp. 1666–1671, July 2007.

[72] A. Kerber, K. Maitra, A. Majumdar, M. Hargrove, R. Carter, and E. Cartier, "Characterization of fast relaxation during bti stress in conventional and advanced cmos devices with HfO_2/TiN gate stacks," *Electron Devices, IEEE Transactions on*, vol. 55, no. 11, pp. 3175–3183, Nov. 2008.

[73] G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, and G. Ghibaudo, "Review on high-k dielectrics reliability issues," *Device and Materials Reliability, IEEE Transactions on*, vol. 5, no. 1, pp. 5–19, March 2005.

[74] A. Kerber, E. Cartier, B. Linder, S. Krishnan, and T. Nigam, "Tddb failure distribution of metal gate/high-k cmos devices on soi substrates," in *Reliability Physics Symposium, IEEE International*, April 2009, pp. 505–509.

[75] T. Nigam, A. Kerber, and P. Peumans, "Accurate model for time-dependent dielectric breakdown of high-k metal gate stacks," in *Reliability Physics Symposium, 2009 IEEE International*, April 2009, pp. 523–530.

[76] J. Di Sarro, Y. Yang, K. Chatty, R. Gauthier, A. Ille, S. Mitra, J. Li, C. Russ, E. Rosenbaum, and D. E. Ioannou, "ESD time-domain characterization of High-k gate dielectric in a 32 nm CMOS technology," in *ESD/EOS Symp.*, 2009.

[77] C.-H. Chen, H.-L. Hwang, and F.-C. Chiu, "Reliability characterization of stress-induced charge trapping in HfO_2 by electrostatic discharge impulse stresses," *Journal of Applied Physics*, vol. 105, no. 10, p. 103910, 2009.

[78] S. Joshi and E. Rosenbaum, "Transmission line pulsed waveform shaping with microwave filters," in *EOS/ESD Symp.*, Sept. 2003, pp. 1–8.

[79] J. Sune and E. Wu, "Statistics of successive breakdown events in gate oxides," *Electron Device Letters, IEEE*, vol. 24, no. 4, pp. 272–274, April 2003.

[80] J. Sune, E. Wu, and S. Tous, "Failure-current based oxide reliability assessment methodology," in *Reliability Physics Symposium, IEEE International*, 27 2008-May 1 2008, pp. 230–239.

[81] N. Raghavan, K. L. Pey, and X. Li, "Detection of high-kappa and interfacial layer breakdown using the tunneling mechanism in a dual layer dielectric stack," *Applied Physics Letters*, vol. 95, no. 22, p. 222903, 2009.

[82] E. Y. Wu and J. Su, "Power-law voltage acceleration: A key element for ultra-thin gate oxide reliability," *Microelectronics and Reliability*, vol. 45, no. 12, pp. 1809 – 1834, 2005.

[83] J. Lee and E. Rosenbaum, "Voltage clamping requirements for ESD protection of inputs in 90nm CMOS technology," in *EOS/ESD Symp.*, 2008, pp. 50–58.

[84] N. Jack and E. Rosenbaum, "Esd protection for high-speed receiver circuits," in *Reliability Physics Symposium, IEEE International*, 2010.

[85] S. Zafar, A. Kumar, E. Gusev, and E. Cartier, "Threshold voltage instabilities in High-κ gate dielectric stacks," *Device and Materials Reliability, IEEE Transactions on*, vol. 5, no. 1, pp. 45–64, Mar. 2005.

[86] D. P. Ioannou, S. Mittl, and G. L. Rosa, "Positive bias temperature instability effects in nMOSFETs with HfO_2/TiN gate stacks," *Device and Materials Reliability, IEEE Transactions on*, vol. 9, no. 2, pp. 128–134, Jun. 2009.

[87] D. Heh, C. Young, and G. Bersuker, "Experimental evidence of the fast and slow charge trapping/detrapping processes in high- k dielectrics subjected to pbti stress," *Electron Device Letters, IEEE*, vol. 29, no. 2, pp. 180–182, Feb. 2008.

[88] A. Kerber, E. Cartier, G. Groeseneken, H. E. Maes, and U. Schwalke, "Stress induced charge trapping effects in SiO_2/Al_2O_3 gate stacks with tin electrodes," *Journal of Applied Physics*, vol. 94, no. 10, pp. 6627–6630, 2003.

[89] Z. Xu, M. Houssa, R. Carter, M. Naili, S. D. Gendt, and M. Heyns, "Constant voltage stress induced degradation in HfO_2/SiO_2 gate dielectric stacks," *Journal of Applied Physics*, vol. 91, no. 12, pp. 10 127–10 129, 2002.

[90] M. Houssa, V. V. Afanasev, A. Stesmans, and M. M. Heyns, "Polarity dependence of defect generation in ultrathin ZrO_2/SiO_2 gate dielectric stacks," *Applied Physics Letter*, vol. 79, p. 3134, 2001.

[91] G. Groeseneken, H. E. Maes, N. Beltran, and R. F. DeKeersmaecker, "A reliable approach to charge-pumping measurements in MOS transistors," *Electron Devices, IEEE Transactions on*, vol. 31, no. 1, pp. 42–53, Jan. 1984.

[92] D. Heh, C. D. Young, G. A. Brown, P. Y. Hung, A. Diebold, G. Bersuker, E. M. Vogel, and J. B. Bernstein, "Spatial distributions of trapping centers in HfO_2/SiO_2 gate stacks," *Applied Physics Letters*, vol. 88, no. 15, p. 152907, 2006.

[93] C. R. Helms and E. H. Poindexter, "The silicon-silicon dioxide system: Its microstructure and imperfections," *Reports on Progress in Physics*, vol. 57, no. 8, pp. 791–852, 1994.

[94] D. J. DiMaria and J. H. Stathis, "Explanation for the oxide thickness dependence of breakdown characteristics of metal-oxide-semiconductor structures," *Applied Physics Letters*, vol. 70, no. 20, pp. 2708–2710, 1997.

[95] N. Sa, J. Kang, H. Yang, X. Liu, Y. He, R. Han, C. Ren, H. Yu, D. Chan, and D.-L. Kwong, "Mechanism of positive-bias temperature instability in sub-1-nm tan/hfn/hfo2 gate stack with low preexisting traps," *Electron Device Letters, IEEE*, vol. 26, no. 9, pp. 610–612, Sept. 2005.

[96] D. A. Buchanan, A. D. Marwick, D. J. DiMaria, and L. Dori, "Hot-electron-induced hydrogen redistribution and defect generation in metal-oxide-semiconductor capacitors," *Journal of Applied Physics*, vol. 76, no. 6, pp. 3595–3608, 1994.

[97] D. J. DiMaria, E. Cartier, and D. A. Buchanan, "Anode hole injection and trapping in silicon dioxide," *Journal of Applied Physics*, vol. 80, no. 1, pp. 304–317, 1996.

[98] Y. Yang, J. Di Sarro, R. J. Gauthier, K. Chatty, J. Li, R. Mishra, S. Mitra, and D. E. Ioannou, "Characterization of high-k/metal gate stack breakdown in the time scale of ESD events," in *Reliability Physics Symposium, IEEE International*, 2010.

[99] R. O'Connor, L. Pantisano, R. Degraeve, T. Kauerauf, B. Kaczer, P. Roussel, and G. Groeseneken, "SILC defect generation spectroscopy in HfSiON using constant voltage stress and substrate hot electron injection," in *Reliability Physics Symposium, IEEE International*, 2008, pp. 324–329.

[100] Y. Huh, M. Lee, J. Lee, H. Jung, T. Li, D. Song, Y. Lee, J. Hwang, Y. Sung, and S. Kang, "A study of ESD-induced latent damage in CMOS integrated circuits," in *Reliability Physics Symposium, IEEE International*, 1998, pp. 279–283.

[101] R. Mishra, S. Mitra, R. Gauthier, D. Ioannou, D. Kontos, K. Chatty, C. Seguin, and R. Halbach, "On the interaction of ESD, NBTI and HCI in 65nm technology," in *Proc. Int. Reliab. Phys. Symp.*, April 2007, pp. 17–22.

[102] R. Arnold and D. Zoroglu, "A quantitative study of emitter ballasting," *Electron Devices, IEEE Transactions on*, vol. 21, no. 7, pp. 385 – 391, Jul 1974.

[103] D. Greenlaw, G. Burbach, T. Feudel, F. Feustel, K. Frohberg, F. Graetsch, G. Grasshoff, C. Hartig, T. Heller, K. Hempel, M. Horstmann, P. Huebler, R. Kirsch, S. Kruegel, E. Langer, A. Pawlowitsch, H. Ruelke, H. Schuehrer, R. Stephan, A. Wei, T. Werner, K. Wieczorek, and M. Raab, "Taking SOI substrates and low-k dielectrics into high-volume microprocessor production," in *Electron Devices Meeting, IEEE International*, dec. 2003, pp. 277–280.

[104] C. Lombardi, S. Manzini, A. Saporito, and M. Vanzi, "A physically based mobility model for numerical simulation of nonplanar devices," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 7, no. 11, pp. 1164–1171, nov 1988.

[105] D. Klaassen, "A unified mobility model for device simulation–I. model equations and concentration dependence," *Solid-State Electronics*, vol. 35, no. 7, pp. 953 – 959, 1992.

[106] S. Sze, *Physics of semiconductor devices, 2nd ed.* John Wiley & Sons, 1981.

[107] B. J. Ballga, *Power semiconductor devices.* PWS Publishing, 1996.

[108] J. P. Colinge, *Silicon-on-insulator technology materials to VLSI, 3rd ed.* Kluwer Academic Publisher, 2004.

[109] J. Di Sarro, K. Chatty, R. Gauthier, and E. Rosenbaum, "Study of design factors affecting turn-on time of silicon controlled rectifiers (SCRS) in 90 and 65nm bulk CMOS technologies," in *Reliability Physics Symposium, IEEE International*, March 2006, pp. 163–168.

[110] S. Cao, S. Beebe, A. Salman, M. Pelella, J.-H. Chun, and R. Dutton, "Field effect diode for effective cdm esd protection in 45 nm soi technology," in *Reliability Physics Symposium, IEEE International*, April 2009, pp. 594–601.

[111] K. Pey, N. Raghavan, X. Li, W. Liu, K. Shubhakar, X. Wu, and M. Bosman, "New insight into the TDDB and post breadkown reliability of novel high-k gate dielectric stacks," in *Reliability Physics Symposium, IEEE International*, 2010, pp. 354–363.

[112] G. Ribes, P. Mora, F. Monsieur, M. Rafik, F. Guarin, G. Yang, D. Roy, W. Chang, and J. Stathis, "High-k gate breakdown statistics modeled by correlated interfacial layer and high-k breakdown path," in *Reliability Physics Symposium, IEEE International*, 2010, pp. 364–368.

[113] M. Radosavljevic, B. Chu-Kung, S. Corcoran, G. Dewey, M. Hudait, J. Fastenau, J. Kavalieros, W. Liu, D. Lubyshev, M. Metz, K. Millard, N. Mukherjee, W. Rachmady, U. Shah, and R. Chau, "Advanced high-k gate dielectric for high-performance short-channel in0.7ga0.3as quantum well field effect transistors on silicon substrate for low power logic applications," in *Electron Devices Meeting, IEEE International*, 2009.

[114] D. Lin, G. Brammertz, S. Sioncke, C. Fleischmann, A. Delabie, K. Martens, H. Bender, T. Conard, W. Tseng, J. Lin, W. Wang, K. Temst, A. Vatomme, J. Mitard, M. Caymax, M. Meuris, M. Heyns, and T. Hoffmann, "Enabling the high-performance InGaAs/Ge cmos: a common gate stack solution," in *Electron Devices Meeting, IEEE International*, 7-9 2009.

Curriculum Vitae

Yang Yang received the B.Sc. (2004) in electrical engineering from Beihang University, Beijing, China and the M.Sc. (2005) in electronics from Queen's University Belfast, Belfast, UK. While at George Mason University, VA, USA, he worked on the various issues related to ESD protection in nano-scale CMOS, including the design & optimization of field-effect-diode (FED) for ESD protection in SOI technology and high-k gate dielectric reliability under ESD-like stress. From Jan. 2009 to Dec. 2009, he was at IBM Microelectronics-System and Technology Group, SRDC, on a student internship in the ESD/Latchup Development Group.